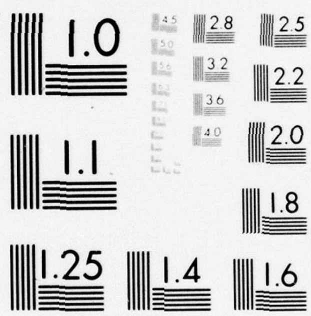
AD-A042 592 CORNELL UNIV ITHACA N Y SCHOOL OF OPERATIONS RESEARC--ETC F/G 5/3
OPTIMAL ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS --ETC(U)
MAY 77 M K VICKERS DAAG29-73-C-0008

UNCLASSIFIED TR-334 NL

1 OF 2
ADA042592



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A042592

SCHOOL
OF
OPERATIONS RESEARCH
AND
INDUSTRIAL ENGINEERING

COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 334

May, 1977

149p.

TR-334

OPTIMAL ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS OF PARAMETERS OF SOME ECONOMETRIC MODELS

by

Mary Kathleen Vickers
Cornell University

Research supported in part by contracts

DAAG29-73-C-0008, U.S. Army Research Office--Durham

N00014-75-C-0586, Office of Naval Research

409 869

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

BIOGRAPHICAL SKETCH

The author was born on December 7, 1949 in Beaumont, Texas, to Leroy and Sylvia Cornwall Vickers. She entered Rice University in September, 1968, as a mathematical science major. There she received the B.A. degree, summa cum laude, in May, 1972. In September, 1972 she began graduate study in operations research at Cornell University. She received the M.S. degree there in June, 1975. Following the completion of her dissertation, she was employed at IBM Federal Systems Division in Owego, New York.

The author is a member of Phi Beta Kappa and the American Statistical Association.

DEDICATION

To my parents, and to Carl

ACKNOWLEDGEMENTS

Professor Lionel Weiss suggested the topic of this thesis, and directed the research there of. I am grateful to him for carefully reading and commenting on several drafts of this work. I would also like to thank Professors William Lucas and Jack Kiefer for serving on my committee. Professor Thomas Santner also read the manuscript and offered valuable comments. Professor Robert Bechhofer has been an informal advisor during my graduate study. I thank him for his continuing interest in my academic and professional work, and for the research and teaching assistantships which supported this work.

My dear friend and colleague Carl Diegert offered much encouragement during this final phase of my graduate study. I am grateful also for his editorial assistance with parts of the text.

During the first, second and sixth semesters of my graduate study, I received fellowships from Cornell University. During the remaining semesters, my work was supported by teaching assistantships and grants from the U.S. Army Research Office, the Office of Naval Research, and the National Science Foundation.

I would like to express my thanks to Vicki Miller and Gail Gunderson for their prompt and accurate typing job.

TABLE OF CONTENTS

CHAPTER I: THEORY OF OPTIMAL PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS UNDER GENERAL CONDITIONS

1.1 Introduction

The purpose of this work is to establish optimality properties for the maximum likelihood estimators, MLE, of a wide variety of econometric models, under assumptions which are as weak as possible. Theorems abound in the literature on optimality properties for MLE. The strength of the assumptions, as well as the conclusions, vary. References to this literature follow in Section 1.2. We shall be interested in proving weak consistency, asymptotic normality and efficiency in the maximum probability, MP, sense. Section 1.3 contains references to the literature for these properties. Section 1.4 contains theorems which allow one to prove that the assumptions required for our optimality properties are fulfilled. The theorem of section 1.5 allows us to prove the optimality properties for one set of parameters, then conclude that they hold on an image set, under certain conditions on the map. Several lemmas which will be useful throughout the following chapters are contained in section 1.6.

1.2 References to the Literature

As pointed out by Crowder [1976], proofs of optimality properties for MLE with dependent observations generally follow one of two lines: that of Wald [1949] or that of Cramér [1946]. Wald's proof of strong consistency of the MLE involves the assumption that the observations are

independent and identically distributed, i.i.d., but no differentiability assumptions are required. These assumptions characterize the continuity and integrability of various functions of the observations' joint density. Neither asymptotic normality nor efficiency results of any kind were obtained. However, Wald pointed out that the method of proof could be extended to dependent variables, under certain conditions. Silvey [1961] adopts a similar argument in his proof of weak consistency. The underlying idea of his proof is that

1.1 $$R(n,\underline{\theta},\underline{\theta}^0) = \log L_n(\underline{\theta}) - \log L_n(\underline{\theta}^0)$$

is a martingale, where $L_n(\underline{\theta})$ is the likelihood function of the observations when $\underline{\theta}\in\Theta\subset R^k$ is the unknown parameter value. Since the ML method chooses $\underline{\theta}$ over $\underline{\theta}^*$ if and only if $R(n,\underline{\theta},\underline{\theta}^*) > 0$, conditions are required which imply the probability that $R(n,\underline{\theta}^0,\underline{\theta}) > 0$, for all $\underline{\theta}$ outside any given neighborhood of $\underline{\theta}^0$, goes to zero as $n \to \infty$. These conditions are:

S1. Θ is compact.

S2. $\mathrm{Var}(R(n,\underline{\theta}^0,\underline{\theta}))^{1/2}/E(R(n,\underline{\theta}^0,\underline{\theta}))$ goes to zero as $n \to \infty$, uniformly outside any open neighborhood of $\underline{\theta}^0$, where the variance and expectation are computed under $\underline{\theta}^0$.

S3. A regularity condition which assures that $\log L_n(\underline{\theta})$ is "sufficiently continuous", i.e. that if $\underline{\theta}^*$ and $\underline{\theta}^{**}$ are outside a given neighborhood of $\underline{\theta}^0$, and $|\underline{\theta}^*-\underline{\theta}^{**}|$ is sufficiently small, then $R(n,\underline{\theta}^0,\underline{\theta}^*) > 0$ implies $R(n,\underline{\theta}^0,\underline{\theta}^{**}) > 0$.

Additional assumptions are required for asymptotic normality. Assuming now that $\theta \epsilon R^1$, these conditions are given as:

S4. $\text{Var}(-\partial^2 \log L_n(\theta)/\partial^2\theta|_{\theta^0})^{1/2} / E(-\partial^2 \log L_n(\theta)/\partial^2\theta|_{\theta^0})$ goes to 0 as $n \to \infty$, where the variance and expectation are computed under $\underline{\theta}^0$.

S5. $\partial \log L_n(\theta)/\partial\theta$ is asymptotically normally distributed.

Silvey cites a martingale central limit theorem to guarantee assumption five. Since $-\partial^2 \log L_n(\theta)/\partial^2\theta$ is a submartingale under appropriate uniform convergence conditions, Silvey concludes that assumption four is often fulfilled, since many submartingales behave in this way.

As mentioned above, Cramér's proof of consistency [1946] for i.i.d. random variables may be modified so that it applies to dependent variables. This is the approach taken by Wald [1948] for a scalar unknown θ, where θ is an interior point of a nondegenerate interval, A. He also demonstrates asymptotic efficiency in the sense described below. Let

1.2 $$c_n(\theta^*) = E(\partial^2 \log L_n(\theta)/\partial^2\theta)|_{\theta=\theta^*}$$

where the expectation is taken under θ^*. Wald defines a sequence of estimators, $\{t_n\}$, to be asymptotically efficient in the wide sense if a random sequence $\{u_n\}$ exists, such that

$$\lim_{n\to\infty} E(u_n) = 0 \quad \text{and} \quad \lim_{n\to\infty} E(u_n^2) = 1$$

where the expectation in both terms is computed under θ^0, and

$$(c_n(\theta^0))^{1/2}(t_n-\theta^0) - u_n$$

converges stochastically to zero as $n \to \infty$. The term "wide sense" is used because the asymptotic distribution of $\{t_n\}$ is not necessarily normal.

Under conditions WC1 through WC4 below, the likelihood equation has at least one root which is consistent. Wald shows that any root which is consistent is also asymptotically efficient in the wide sense.

WC1. The first, second and third derivatives of $L_n(\theta)$ exist for all θ in A, with probability one. Furthermore,

$$\operatorname*{lub}_{\theta \in A} \left| \frac{\partial^i L_n(\theta)}{\partial^i \theta} \right| \qquad i=1,2,3$$

is integrable over the observation space.

WC2. For any θ in A,

$$\lim_{n\to\infty} c_n(\theta) = \infty$$

WC3. $\mathrm{Var}(\partial^2 \log L_n(\theta)/\partial^2\theta)^{1/2}/E(\partial^2 \log L_n(\theta)/\partial^2\theta)$ goes to 0 as $n \to \infty$, where the expectation and variance are computed under any θ in A.

WC4. There exists a positive δ such that for any θ in A

$$(c_n(\theta))^{-1} E\left\{ \operatorname*{lub}_{\theta^* \text{ s.t. } |\theta^*-\theta| \le \delta} \left| \frac{\partial^3 \log L_n(\theta)}{\partial^3\theta} \right| \right\}$$

is a bounded function of n, when the expectation is taken under θ.

Bar-Shalom [1971] gives conditions in terms of the density of individual observations, conditional upon all past observations, under which the optimality properties of Wald [1948] hold. Let

$$p_1 \equiv p(y_1;\theta)$$
$$p_k \equiv p(y_k/y_{k-1},\ldots,y_1;\theta) \qquad k = 2,\ldots,n$$

be the conditional density of the k'th observation, y_k, when θ is the value of the unknown scalar parameter, for $k \geq 2$. The unconditional density of y_1 is p_1. As before, assume Θ is the domain of the unknown parameter, and θ^0 is the true value. The following regularity conditions must be satisfied for all k:

B1. $\partial^i \log p_k/\partial^i\theta$ exists for all $\theta \in \Theta$, for $i = 1,2,3$.

B2. $E(\partial \log p_k/\partial\theta|_{\theta^0}) = 0$, where the expectation is taken under θ^0.

B3. $J_k(\theta^0) \equiv E((\partial \log p_k/\partial\theta)^2) \leq C_1 < \infty$, where C_1 is independent of k.

B4. $E(\partial^2 \log p_k/\partial^2\theta)|_{\theta^0} = -J_k(\theta^0)$

B5. There exists a measurable function $H_k(y_1,\ldots,y_k)$ such that

$$\left|\frac{\partial^3 \log p_k}{\partial^3\theta}\right| < H_k(y_1,\ldots,y_k)$$

for all $\theta \in \Theta$, and H_k is finite, except on a set of measure zero.

B6. $\lim_{|k-j|\to\infty} E\left(\frac{\partial \log p_k}{\partial\theta} \frac{\partial \log p_j}{\partial\theta}\right)\Big|_{\theta^0} = 0$

B7. $\mathrm{Var}(\partial^2 \log p_k/\partial^2\theta)\Big|_{\theta^0} \leq C_2 < \infty$, where C_2 is independent of k, and

$$\lim_{|k-j|\to\infty} \mathrm{cov}\left(\frac{\partial^2 \log p_k}{\partial^2\theta}, \frac{\partial^2 \log p_j}{\partial^2\theta}\right)\Big|_{\theta^0} = 0.$$

Under conditions B1 through B7, the MLE of θ^0 is weakly consistent. If condition six is strengthened to

B6'. $E\left(\frac{\partial \log p_k}{\partial\theta} \quad \frac{\partial \log p_j}{\partial\theta}\right) = 0$, for all $j \neq k$.

then the MLE is also wide sense asymptotically efficient.

Bhat [1974] shows that the conditions of Bar-Shalom may be simplified and a martingale central limit theorem invoked to imply consistency and asymptotic normality with asymptotic efficiency in the wide sense. (Wald [1948] refers to the latter property as asymptotic efficiency in the strict sense.) Bhat requires that

BH1. The range of the k'th observation given the previous observations does not depend on θ. The derivatives up to third order of p_k are continuous for $\theta \in \Theta$. Differentiation with respect to θ up to third order of $\log p_k$ may be carried out under the integral, where the integration is with respect to the observations.

BH2. $|\partial^3 \log p_k/\partial^3\theta|$ is bounded in probability uniformly for all $y_1,\ldots,y_k$ and k.

BH3. $\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} J_k(\theta^0) = J(\theta^0) > 0$

BH4. For some $\delta > 0$,

$$n^{-1-\delta/2} \sum_{k=1}^{n} E(|\partial \log p_k/\partial\theta|^{2+\delta}/y_1,\ldots,y_k) \to 0 \qquad \text{a.s.}$$

BH5. $n^{-2} \sum_{k=1}^{n} \mathrm{var}(\partial^2 \log p_k/\partial^2\theta)|_{\theta^0}$ remains finite as $n \to \infty$.

Condition BH1 implies that $\partial \log L_n(\theta)/\partial\theta$ is a martingale. The fact that

$$\sum_{k=1}^{n} [\partial^2 \log p_k/\partial^2\theta + E((\partial \log p_k/\partial\theta)^2/y_1,\ldots,y_{k-1})]$$

is a martingale is also required.

Crowder [1976] also makes use of martingale theory to prove asymptotic normality. His proof of consistency does not involve assumptions about the third derivatives of the log likelihood function. To show consistency, he defines $L_n''(\underline{\theta},\underline{\theta}^0)$ to be the matrix of second derivatives of the log likelihood function, with the rows evaluated at possibly different points on the line segment between $\underline{\theta}$ and $\underline{\theta}^0$.

The two term Taylor expansion of the first derivative of the log likelihood is

$$\text{1.3} \qquad \partial \log L_n(\underline{\theta})/\partial\underline{\theta}|_{\underline{\theta}} = \partial \log L_n(\underline{\theta})/\partial\underline{\theta}|_{\underline{\theta}^0} + L_n''(\underline{\theta},\underline{\theta}^0)(\underline{\theta}-\underline{\theta}^0)$$

For weak consistency of the MLE Crowder required that:

C1. $L_n''(\underline{\theta},\underline{\theta}^0)$ be continuous in $\underline{\theta}$ throughout some neighborhood of $\underline{\theta}^0$.

C2. Uniform convergence conditions apply so that, when the expectation and variance are taken under $\underline{\theta}^0$,

$E(\partial \log L_n(\underline{\theta})/\partial\underline{\theta}|_{\underline{\theta}^0}) = 0$

$\mathrm{Var}(\partial \log L_n(\underline{\theta})/\partial\underline{\theta}|_{\underline{\theta}^0}) = E(-L_n''(\underline{\theta}^0,\underline{\theta}^0)) \equiv B_n$

where B_n is assumed to be positive definite.

C3. There exists $\Delta > 0$ and a positive sequence c_n, $c_n \to \infty$, not functions of $\underline{\theta}$, such that $|\underline{\theta}-\underline{\theta}^0| = \delta \leq \Delta \Rightarrow$

$$P\left[\frac{(\underline{\theta}-\underline{\theta}^0)^T \; B_n^{-1/2} L_n''(\underline{\theta}^0,\underline{\theta})(\underline{\theta}-\underline{\theta}^0)}{c_n^{1/2}} \geq \delta^2\right] \to 1$$

as $n \to \infty$.

Assumption C3 implies that

$$B_n^{-1/2} L_n''(\underline{\theta}^0,\underline{\theta})$$

goes to infinity in some sense as $n \to \infty$; this corresponds to assumption WC2 of Wald.

To prove asymptotic normality, 1.3 is evaluated at the MLE $\hat{\underline{\theta}}(n)$ to obtain

$$\hat{\underline{\theta}}(n) - \underline{\theta}^0 = -L_n''(\underline{\theta}^0,\hat{\underline{\theta}}(n))^{-1} B_n B_n^{-1}(\partial \log L_n(\underline{\theta})/\partial\underline{\theta})|_{\underline{\theta}^0}$$

It follows from assumption C3 that

$$B_n^{-1} L_n''(\underline{\theta}^0,\hat{\underline{\theta}}(n)) \to I_k$$

stochastically, as $n \to \infty$, where I_k is the k by k identity matrix. Crowder utilizes a martingale central limit theorem to establish that

$$B_n^{-1}(\partial \log L_n(\underline{\theta})/\partial\underline{\theta})|_{\underline{\theta}^0}$$

is asymptotically normally distributed.

1.3 Weiss's Theorem for Asymptotic Efficiency in the Maximum Probability Sense

The definition of asymptotic efficiency in the previous section involved the requirement that the variance of the limiting distribution of the estimator attain the Cramér-Rao variance. As Weiss and Wolfowitz [1967] point out, comparing the performance of estimators by their variances is meaningless for many distributions other than normal. Furthermore, restricting the class of estimators to those which are asymptotically normally distributed is not reasonable from a statistical standpoint. Weiss and Wolfowitz [1967] define asymptotic efficiency to overcome these problems. An estimator is said to be asymptotically efficient with respect to any given set containing the true parameter value if it maximizes, among estimators which satisfy a uniformity condition for convergence, the limiting probability of being in the given set. The authors' maximum probability, MP, estimator is asymptotically efficient in this sense; hence the estimators enjoying this property will be said to be (asymptotically) efficient in the MP sense.

To determine conditions under which MLE for various econometric models are asymptotically efficient in the MP sense, Weiss's [1973] theorem will be used. Before stating the assumptions, we give the

notation used by Weiss. $\underline{\theta}^0$ will denote any given vector in the interior of $\Theta \subset R^k$. There must exist $2k$ sequences of nonrandom positive quantities $K_1(n),\ldots,K_k(n)$, $M_1(n),\ldots,M_k(n)$ which satisfy

$$\lim_{n\to\infty} K_i(n) = \infty \qquad \lim_{n\to\infty} M_i(n) = \infty \qquad \lim_{n\to\infty} M_i(n)/K_i(n) = 0 \quad i = 1,\ldots,k.$$

Define the closed set

$$N_n(\underline{\theta}^0) = \{\underline{\theta}\varepsilon\Theta : |\theta_i - \theta_i^0| \leq M_i(n)/K_i(n), \text{ for } i = 1,\ldots,k\}.$$

We now give two assumptions.

WE1. There exist nonrandom continuous functions $B_{ij}(\underline{\theta}^0)$, for $i,j = 1,\ldots,k$, for $\underline{\theta}^0 \varepsilon \Theta$ such that

$$-\frac{1}{K_i(n)K_j(n)} \frac{\partial^2 \log L_n(\underline{\theta})}{\partial\theta_i\,\partial\theta_j}\Bigg|_{\underline{\theta}^0}$$

converges stochastically as $n \to \infty$ to $B_{ij}(\underline{\theta}^0)$, when $\underline{\theta}^0$ is the true parameter value.

Define $\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n) \equiv \dfrac{-1}{K_i(n)K_j(n)} \dfrac{\partial^2 \log L_n(\underline{\theta})}{\partial\theta_i\,\partial\theta_j}\Bigg|_{\underline{\theta}} - B_{ij}(\underline{\theta}^0)$

Let $R_n(\underline{\theta}^0,\gamma)$ denote the set of observations for which

$$\sum_{i=1}^{k}\sum_{j=1}^{k} M_i(n)M_j(n) \sup_{\underline{\theta}\in N_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| < \gamma.$$

The second assumption is

WE2. There exist sequences of nonrandom positive quantities,

$\{\gamma(n,\underline{\theta}^0)\}$, $\{\delta(n,\underline{\theta}^0)\}$ such that $\gamma(n,\underline{\theta}^0) \to 0$, $\delta(n,\underline{\theta}^0) \to 0$

and

$P(R_n(\underline{\theta}^0, \gamma(n,\underline{\theta}^0)) > 1 - \delta(n,\underline{\theta}^0)$

for all $\underline{\theta} \in N_n(\underline{\theta}^0)$ where the probability is computed

under $L_n(\underline{\theta})$.

Under assumptions WE1 and WE2, the MLE of $\underline{\theta}$ are consistent, asymptotically normally distributed and asymptotically efficient in the MP sense. It should be noted that the assumptions do not require i.i.d. observations. Hereafter, sequences of positive quantities which converge to zero will be called null sequences.

1.4 Sufficient Conditions for Asymptotic Efficiency in the MP Sense

To facilitate the use of Weiss's theorem we prove three theorems which imply that Weiss's two assumptions are fulfilled in many econometric models. Thus the MLE for the econometric parameters have the optimality properties we have discussed.

Theorem 1.1 is essentially a weak law of large numbers for the second partial derivatives of the log likelihood function. However, the stochastic convergence is shown to be uniform in any decreasing neighbohood of Θ. (Recall that every point of Θ is an interior point.) Theorems 1.2 and 1.3 imply that Weiss's assumption WE2 holds. The following notation is used for $i,j = 1,\ldots,k$:

$$D_{ij}(n,\underline{\theta}^*) = -n^{-1}(\partial^2 \log L_n(\underline{\theta})/\partial\theta_i\partial\theta_j)\Big|_{\underline{\theta}^*}$$

$B_{ij}(n,\underline{\theta}^*) = E(D_{ij}(n,\underline{\theta}^*))$ where the expectation is taken under $\underline{\theta}^*$.

$B_{ij}(\underline{\theta}) = \lim_{n\to\infty} B_{ij}(n,\underline{\theta})$ if the limit exists.

It is sufficient for all our models that $K_i(n) = n^{1/2}$ for $i = 1,\ldots,k$ and $M_i(n) = M(n)$ for $i = 1,\ldots,k$. Further restrictions on $M(n)$ are given in theorem 1.3.

Theorem 1.1

Let $\{e_i(n)\}$, $i = 1,\ldots,k$ be positive sequences which each go to zero. For any $\underline{\theta}^0$ in Θ, let $A_n(\underline{\theta}^0) = \{\underline{\theta}\varepsilon\Theta\colon |\theta_i-\theta_i^0| < e_i(n),\ i=1,\ldots,k\}$. Assume that there exist two null sequences $\{\nu_{ij}(n,\underline{\theta}^0)\}$ and $\{\eta_{ij}(n,\underline{\theta}^0)\}$ such that, for all $\underline{\theta}$ in $A_n(\underline{\theta}^0)$,

$$|B_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta})| \le \nu_{ij}(n,\underline{\theta}^0) \quad \text{and}$$

$$|Var(D_{ij}(n,\underline{\theta}))| \le \eta_{ij}(n,\underline{\theta}^0)$$

where the variance is taken under $\underline{\theta}$. Then there exist two null sequences $\{\delta_{ij}^1(n,\underline{\theta}^0)\}$ and $\{\delta_{ij}^2(n,\underline{\theta}^0)\}$ such that,

$$P\{|D_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta})| \ge \delta_{ij}^1(n,\underline{\theta}^0)\} \le \delta_{ij}^2(n,\underline{\theta}^0) \quad \text{for all } \underline{\theta}\in A_n(\underline{\theta}^0)$$

where the probability is taken under $\underline{\theta}$. Furthermore, if $B_{ij}(n,\underline{\theta})$ is continuous as a function of $\underline{\theta}$, then $B_{ij}(\underline{\theta})$ is also continuous.

Proof:

For any $\epsilon_n > 0$, Chebycheff's inequality implies

$$P\{|D_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta})| \geq \epsilon_n\}$$

$$\leq \epsilon_n^{-2}(\mathrm{Var}(D_{ij}(n,\underline{\theta})) + (B_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta}))^2)$$

$$\leq \epsilon_n^{-2}(\eta_{ij}(n,\underline{\theta}^0) + \nu_{ij}(n,\underline{\theta}^0)) \quad \text{for all } \underline{\theta}\epsilon A_n(\underline{\theta}^0)$$

Let $\epsilon_n = \max\{\eta_{ij}(n,\underline{\theta}^0)^{1/4}, \nu_{ij}(n,\underline{\theta}^0)^{1/4}\}$

Then $\delta_{ij}^2(n,\underline{\theta}^0) = \epsilon_n^{-2}(\eta_{ij}(n,\underline{\theta}^0) + \nu_{ij}(n,\underline{\theta}^0))$

$$\leq (\eta_{ij}(n,\underline{\theta}^0))^{1/2} + (\nu_{ij}(n,\underline{\theta}^0))^{1/2}$$

is a null sequence. Furthermore, $\delta_{ij}^1(n,\underline{\theta}^0) = \epsilon_n$ is also a null sequence, and the first claim is proved. To prove the continuity of $B_{ij}(\underline{\theta})$, we note first that the continuity of $B_{ij}(n,\underline{\theta})$ implies the existence of a null sequence $\{\alpha_n(\underline{\theta}^0)\}$ such that if $\underline{\theta},\underline{\theta}'$ in $A_n(\underline{\theta}^0)$,

$$|B_{ij}(n,\underline{\theta}) - B_{ij}(n,\underline{\theta}')| \leq \alpha_n(\underline{\theta}^0)$$

For any $\epsilon > 0$, choose n so large that $2\nu_{ij}(n,\underline{\theta}^0) + \alpha_n(\underline{\theta}^0) < \epsilon$: Then, for $\underline{\theta},\underline{\theta}'$ in $A_n(\underline{\theta}^0)$,

$$|B_{ij}(\underline{\theta}) - B_{ij}(\underline{\theta}')| \leq |B_{ij}(\underline{\theta}) - B_{ij}(n,\underline{\theta})| + |B_{ij}(n,\underline{\theta}) - B_{ij}(n,\underline{\theta}')|$$

$$+ |B_{ij}(n,\underline{\theta}') - B_{ij}(\underline{\theta}')|$$

Therefore continuity of B_{ij} is proven.

Theorem 1.2

For $A_n(\underline{\theta}^0)$ as defined in theorem 1.1, suppose the null sequences

$\{\delta^1_{ij}(n,\underline{\theta}^0)\}$ and $\{\delta^2_{ij}(n,\underline{\theta}^0)\}$ of theorem 1.1 exist, and that $B_{ij}(\underline{\theta})$ is continuous in Θ. Then there exist two null sequences $\{\gamma^1_{ij}(n,\underline{\theta}^0)\}$ and $\{\gamma^2_{ij}(n,\underline{\theta}^0)\}$ such that

$$P\left[\sup_{\underline{\theta}\in A_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \geq \gamma^1_{ij}(n;\underline{\theta}^0)\right] \leq \gamma^2_{ij}(n,\underline{\theta}^0)$$

where the probability is taken under $\underline{\theta}$, for all $\underline{\theta}\in A_n(\underline{\theta}^0)$.

Proof:

We can express

$$\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n) = D_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta}) + B_{ij}(\underline{\theta}) - B_{ij}(\underline{\theta}^0)$$

For all $\underline{\theta}\in A_n(\underline{\theta}^0)$,

$$|D_{ij}(n,\underline{\theta}) - B_{ij}(\underline{\theta})| < \delta^1_{ij}(n,\underline{\theta}^0)$$

with probability (under $\underline{\theta}$) greater than $1 - \delta^2_{ij}(n,\underline{\theta}^0)$. The continuity of B_{ij} implies that there exists a null sequence $\{\rho_{ij}(n,\underline{\theta}^0)\}$ such that whenever $\underline{\theta}$ is in $A_n(\underline{\theta}^0)$,

$$|B_{ij}(\underline{\theta}) - B_{ij}(\underline{\theta}^0)| \leq \rho_{ij}(n,\underline{\theta}^0)$$

Let $\gamma^1_{ij}(n,\underline{\theta}^0) = \delta^1_{ij}(n,\underline{\theta}^0) + \rho_{ij}(n,\underline{\theta}^0)$ and $\gamma^2_{ij}(n,\underline{\theta}^0) = \delta^2_{ij}(n,\underline{\theta}^0)$.

Then $|\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \leq \gamma^1_{ij}(n,\underline{\theta}^0)$ for all $\underline{\theta}$ in $A_n(\underline{\theta}^0)$ implies

$$\sup_{\underline{\theta} \in A_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n) \le \gamma^1_{ij}(n,\underline{\theta}^0)$$

and the probability under $\underline{\theta}$ of the first event is at least $1-\gamma^2_{ij}(n,\underline{\theta}^0)$ for all $\underline{\theta}$ in $A_n(\underline{\theta}^0)$. Therefore the theorem is proven.

Theorem 1.3

Assume the existence of null sequences $\{\gamma^1_{ij}(n,\underline{\theta}^0)\}$ and $\{\gamma^2_{ij}(n,\underline{\theta}^0)\}$ for $i,j = 1,\ldots,k$ of theorem 1.2. Then there exist two null sequences $\{\gamma^1(n,\underline{\theta}^0)\}$ and $\{\gamma^2(n,\underline{\theta}^0)\}$ and a positive sequence $\{M(n)\}$ such that $M(n) \to \infty$, $n^{-1/2} M(n) \to 0$, and

$$P\left[\sum_{i=1}^{k} \sum_{j=1}^{k} M^2(n) \sup_{\underline{\theta} \in N_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \ge \gamma^1(n,\underline{\theta}^0)\right] \le \gamma^2(n,\underline{\theta}^0)$$

where the probability is computed under $\underline{\theta}$, for every $\underline{\theta}$ in $N_n(\underline{\theta}^0)$.

Proof:

Let $M(n) = \min\{n^{1/2-\delta}, \min_{1 \le i,\, j \le k} \{\gamma^1_{ij}(n,\underline{\theta}^0)^{-1/4}\}\}$ where $0 < \delta < 1/2$.

Then $M(n) \to 0$, $K(n)^{-1}$ $M(n) = n^{-1/2}$ $n^{1/2-\delta} \to 0$ as $n \to \infty$.

By assumption,

$$P\left[M^2(n) \sup_{\underline{\theta} \in N_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \ge M^2(n)\gamma^1_{ij}(n,\underline{\theta}^0)\right] \le \gamma^2_{ij}(n,\underline{\theta}^0)$$

where the probability is taken under $\underline{\theta}$, for all $\underline{\theta}$ in $N_n(\underline{\theta}^0)$.

Let $\gamma^1(n,\underline{\theta}^0) \equiv \sum_{i=1}^{k} \sum_{j=1}^{k} M^2(n)\gamma^1_{ij}(n,\underline{\theta}^0) \le \sum_{i=1}^{k} \sum_{j=1}^{k} (\gamma^1_{ij}(n,\underline{\theta}^0))^{1/2}$ so that $\{\gamma^1(n,\underline{\theta}^0)\}$ is a null sequence. Let

$$\gamma^2(n,\underline{\theta}^0) = \sum_{i=1}^{k} \sum_{j=1}^{k} \gamma^2_{ij}(n,\underline{\theta}^0)$$

$\{\gamma^2(n,\underline{\theta}^0)\}$ is clearly a null sequence.

Let E_{ij} be the event that $\{M^2(n) \sup_{\underline{\theta}\in N_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \geq M^2(n)\gamma^1_{ij}(n,\underline{\theta}^0)\}$

Then

$$P\left[\sum_{i=1}^{k}\sum_{j=1}^{k} M^2(n) \sup_{\underline{\theta}\in N_n(\underline{\theta}^0)} |\epsilon_{ij}(\underline{\theta},\underline{\theta}^0,n)| \geq \gamma^1(n,\underline{\theta}^0)\right] \leq$$

$$P\left(\bigcup_{i,j=1}^{k} E_{ij}\right) \leq \sum_{i=1}^{k}\sum_{j=1}^{k} P(E_{ij}) \leq \sum_{i=1}^{k}\sum_{j=1}^{k}\gamma^2_{ij}(n,\underline{\theta}^0) \equiv \gamma^2(n,\underline{\theta}^0)$$

The result is proved.

We note that to apply theorems 1.1, 1.2 and 1.3, it is necessary only to verify that, for all $i,j = 1,\ldots,k$, $B_{ij}(n,\underline{\theta})$ is continuous as a function of $\underline{\theta}$, $B_{ij}(n,\underline{\theta})$ converges as $n \to \infty$ uniformly in a compact set containing $\underline{\theta}^0$, and that the variance of $D_{ij}(n,\underline{\theta})$, computed under $\underline{\theta}$, converges uniformly to zero in a compact set containing $\underline{\theta}^0$.

1.5 Optimality Results for Functions of the Model Parameters

For some models, it is convenient to prove optimality properties for a parameter set, then infer that the same properties hold for an image set of the original parameter set. This is convenient, for instance, when the error variables are an autoregressive process. It is also convenient when the model is a linear system of equations. Nocturne [1970] has proved a theorem which justifies this inference when the observations form a Markov process. We will prove a similar theorem. Let the mapping which defines the image parameter set be

1.4 $$H: (\theta_1,\ldots,\theta_k) \to (H_1(\underline{\theta}),\ldots,H_m(\underline{\theta}))$$

where $m \leq k$. We want to prove that optimality properties which hold for $\underline{\theta}$ also hold for $\underline{H}$.

Theorem 1.4.

Suppose a parameter map H is defined by 1.4. If $m = k$, the map is one-to-one. If $m < k$, then there exist $k - m$ functions $H_{m+1}, \ldots, H_k$ such that the map $\underline{\theta} \leftrightarrow \underline{H}$ is one-to-one. Assume that the hypotheses for theorems 1.1 and 1.2 hold for $D_{ij}(n,\underline{\theta})$, for $i,j = 1,\ldots,k$. Assume also that

$$\frac{\partial \theta_p}{\partial H_i}, \quad \frac{\partial^2 \theta_p}{\partial H_i \partial H_j}$$

exist and are continuous functions of $\underline{H}$, for $p,i,j = 1,\ldots,k$. Then $\underline{H}(\hat{\underline{\theta}}(n))$ is the MLE of $\underline{H}(\underline{\theta})$, where $\hat{\underline{\theta}}(n)$ is the MLE of $\underline{\theta}$. Furthermore, $\underline{H}(\hat{\underline{\theta}}(n))$ is consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof: Since $\underline{H} = (H_1,\ldots,H_k)$ is bijective, the joint density of the observations may be written as a function of $\underline{\theta}$ or as a function of $\underline{H}$, viz.

$$L_n(\underline{\theta}) \equiv \overline{L}_n(\underline{H})$$

It is clear that $\underline{H}(\hat{\underline{\theta}}(n))$ is the MLE of $\underline{H}(\underline{\theta})$. To verify the hypotheses of theorems 1.1, 1.2 and 1.3 for

$$\overline{D}_{ij}(n,\underline{H}) = -n^{-1} \frac{\partial^2 \log \overline{L}_n(\underline{H})}{\partial H_i \partial H_j} ,$$

we note that

$$\bar{D}_{ij}(n,\underline{H}) = \sum_{a=1}^{k} \frac{\partial^2 \theta_a}{\partial H_i \partial H_j} \left[-n^{-1} \frac{\partial \log L_n(\underline{\theta})}{\partial \theta_a} \right] \tag{1.5}$$

$$+ \sum_{a=1}^{k} \sum_{b=1}^{k} \frac{\partial \theta_a}{\partial H_i} \frac{\partial \theta_b}{\partial H_j} \left[-n^{-1} \frac{\partial^2 \log L_n(\underline{\theta})}{\partial \theta_a \, \partial \theta_b} \right]$$

The expected value, under $\underline{H}$, of 1.5 is

$$\bar{B}_{ij}(n,\underline{H}) = \sum_{a=1}^{k} \sum_{b=1}^{k} (\partial \theta_a / \partial H_i)\,(\partial \theta_b / \partial H_j)\, B_{ij}(n,\underline{\theta}).$$

Since the hypotheses of theorem 1.1 hold for $B_{ij}(n,\underline{\theta})$ and the first partial derivatives of $\underline{\theta}_a$ are continuous as functions of $\underline{\theta}$, $\bar{B}_{ij}(n,\underline{H})$ is a continuous function of $\underline{H}$, and converges, as $n \to \infty$, to $\bar{B}_{ij}(\underline{H})$. Furthermore, the convergence is uniform in decreasing neighborhoods of $\underline{H}(\underline{\theta}^0)$. The variance of $\bar{D}_{ij}(n,\underline{H})$ is

$$\sum_{a=1}^{k} \sum_{b=1}^{k} \frac{\partial^2 \theta_a}{\partial H_i \partial H_j} \frac{\partial^2 \theta_b}{\partial H_i \partial H_j} \operatorname{cov} \left(-n^{-1} \frac{\partial \log L_n(\underline{\theta})}{\partial \theta_a} , -n^{-1} \frac{\partial \log L_n(\underline{\theta})}{\partial \theta_b} \right) \tag{1.6}$$

$$+ \sum_{a=1}^{k} \sum_{b=1}^{k} \sum_{c=1}^{k} \sum_{d=1}^{k} \frac{\partial \theta_a}{\partial H_i} \frac{\partial \theta_b}{\partial H_j} \frac{\partial \theta_c}{\partial H_i} \frac{\partial \theta_d}{\partial H_j} \operatorname{cov}(D_{ab}(n,\underline{\theta}),\ D_{cd}(n,\underline{\theta}))$$

In the first term of 1.6, the covariance term is less than or equal to $\mathrm{Var}(-n^{-1} \partial \log L_n(\underline{\theta})/\partial \theta_a)^{1/2}\ \mathrm{Var}(-n^{-1} \partial \log L_n(\underline{\theta})/\partial \theta_b)^{1/2}$.

Under the appropriate uniform convergence theorems, we have

$$\mathrm{Var}(-n^{-1} \partial \log L_n(\underline{\theta})/\partial \theta_a) = n^{-1} B_{aa}(n,\underline{\theta})$$

Since the second partial derivatives of θ_a are continuous as functions of $\underline{\theta}$ and $B_{aa}(n,\underline{\theta})$ converges as $n \to \infty$, uniformly in a decreasing neighborhood of $\underline{\theta}^0$, the first term of 1.6 goes to zero as $n \to \infty$, uniformly in a decreasing neighborhood of $\underline{H}(\underline{\theta}^0)$.

In the second term of 1.6, the absolute value of the covariance term is bounded by

$$\mathrm{Var}(D_{ab}(n,\underline{\theta}))^{1/2}\ \mathrm{Var}(D_{cd}(n,\underline{\theta}))^{1/2}$$

By assumption, both factors converge to zero as $n \to \infty$, uniformly in a decreasing neighborhood of $\underline{\theta}^0$. Since the first partial derivatives of $\underline{\theta}$ with respect to $\underline{H}$ are continuous functions of $\underline{\theta}$, the second term of 1.6 goes to zero as required, uniformly in a decreasing neighborhood of $\underline{H}(\underline{\theta}^0)$.

We have shown that the hypotheses of theorems 1.1, 1.2 and 1.3 hold for $\bar{D}_{ij}(n,\underline{H})$ for $i,j=1,\ldots,k$. Therefore $\underline{H}(\hat{\underline{\theta}}(n))$ is consistent, asymptotically normally distributed and efficient in the MP sense. The theorem is proved.

1.6 Four Useful Lemmas

This section contains four lemmas which are required in the following chapters. These lemmas establish the convergence of certain series which will appear frequently. The proofs of these lemmas are simple, and are therefore not given.

Lemma 1.1 If $|p| < 1$, and $\delta > 0$, then

$$\lim_{n\to\infty} n^{-1-\delta} \sum_{t=1}^{n} \sum_{s=1}^{n} p^{|t-s|} = 0$$

<u>Lemma 1.2</u> If $|p| < 1$, and $\delta > 0$, then

$$\lim_{n\to\infty} n^{-1-\delta} \sum_{t=1}^{n} \sum_{s=1}^{n} |t-s| p^{|t-s|} = 0$$

<u>Lemma 1.3</u> For $t = 1,2,\ldots$, let $S_t(\underline{\theta})$ be a continuous function of $\underline{\theta}$ over a compact set C. Let $T_n(\underline{\theta}) = n^{-1} \sum_{t=1}^{n} S_t(\underline{\theta})$. If

$$\lim_{t\to\infty} S_t(\underline{\theta})$$

exists, then

$$\lim_{n\to\infty} T_n(\underline{\theta})$$

exists. Also, convergence of $T_n(\underline{\theta})$ as $n \to \infty$ is uniform in C if convergence of S_t as $t \to \infty$ is uniform in C.

<u>Lemma 1.4</u> For $k,m = 1,2,\ldots$, let $S_k(\underline{\theta})$, $T_m(\underline{\theta})$ be continuous functions of $\underline{\theta}$ over a compact set C. Suppose

$$\lim_{t\to\infty} \sum_{k=1}^{t} |S_k(\underline{\theta})| \equiv S(\underline{\theta}) < \infty \qquad \text{for every } \underline{\theta} \text{ in } C.$$

and

$$\lim_{t\to\infty} \sum_{m=1}^{t} |T_m(\underline{\theta})| \equiv T(\underline{\theta}) < \infty \qquad \text{for every } \underline{\theta} \text{ in } C.$$

Then,

$$\lim_{t\to\infty} \sum_{k=1}^{t} \sum_{m=1}^{t} S_k(\underline{\theta})\, T_m(\underline{\theta}) u_{km}$$

exists for every $\underline{\theta}$ in C if there exists an $M > 0$ such that $|u_{km}| < M$ for all k,m. The convergence of the last term is uniform over C if the convergence to $S(\underline{\theta})$, $T(\underline{\theta})$ is uniform over C.

CHAPTER II: LINEAR MODELS WITH LAGGED DEPENDENT VARIABLES AS REGRESSORS

2.1 Results from the Literature

It is desired to determine optimal properties for the MLE of the single equation model

2.1 $$y_t = \sum_{g=1}^{G} \beta_g y_{t-g} + \sum_{h=1}^{K} \beta_{G+h} x_{t,h} + \varepsilon_t \qquad \text{for } t=1,2,\ldots$$

where $\{\varepsilon_t, t=1,2,\ldots\}$ are independently, normally distributed, with zero expectation and unknown variance σ^2. In all results and theorems in this section, the stability condition will be assumed to hold; i.e., the roots of the polynomial equation

2.2 $$\rho^G - \beta_1 \rho^{G-1} - \ldots - \beta_G = 0$$

are less than one in absolute value. This assumption assures that the oscillations of the error variables and exogenous variables are not amplified in the $\{y_t, t=1,2,\ldots\}$.

Koopmans, Rubin, and Leipnik [1950] assume that the variables $\{x_{t,h}\}$ are nonrandom and that

$$n^{-1} \sum_{t=1}^{n} x_{t,h} x_{t+c,k}$$

converges for h,k in $\{1,\ldots,H\}$ and for any positive integer c.

Under these assumptions the authors show that the MLE of $\underline{\beta} = (\beta_1, \beta_2, \ldots, \beta_K)$ are consistent and that their asymptotic distribution is jointly normal with mean vector $\underline{0}$ and covariance matrix $\sigma^2 N^{-1}$ where the matrix N may be written as

2.3 $$N = \begin{bmatrix} M_{yy} & M_{y0} \\ M_{y0}^T & M_{00} \end{bmatrix}.$$

M_{yy} is the $G \times G$ matrix whose h,k'th element is the stochastic limit of

$$n^{-1} \sum_{t=1}^{n} y_{t-h} y_{t-k} \qquad h,k = 1, \ldots, G.$$

M_{y0} is the $G \times K$ matrix whose h,k'th element is the stochastic limit of

$$n^{-1} \sum_{t=1}^{n} y_{t-h} x_{t,k} \qquad h = 1, \ldots, G;\ k = 1, \ldots, K$$

and M_{00} is the $K \times K$ matrix whose h,k'th element is the limit of

$$n^{-1} \sum_{t=1}^{n} x_{t,h} x_{t,k} \qquad h,k = 1, \ldots, K.$$

The MLE, $\hat{\underline{\beta}}(n)$, of $\underline{\beta}$ are shown to be efficient in the sense that the asymptotic concentration ellipsoid of $n^{1/2}(\hat{\underline{\beta}}(n) - \underline{\beta})$, which is the set $\{\underline{e} : \underline{e}^T N \underline{e} \leq \sigma^2\}$, is contained in that of every other consistent estimator. Normality of $\{\varepsilon_t\}$ is not necessary for the proofs of consistency and asymptotic normality of the MLE. This condition may be relaxed to the condition that the error distributions have zero means and finite moments of every order. However, the normality

assumption is required for the efficiency result.

Theil [1971] also considers the linear model with lagged dependent variables. For consistency and asymptotic normality of the estimators, the stability condition 2.2 is assumed. The convergence of

$$(n-c)^{-1} \sum_{t=1}^{n-c} x_{t,h} x_{t+c,k} \qquad h,k=1,\ldots,K$$

is required, for $c=0,1,\ldots,G$. (Koopmans, Rubin, and Leipnik [1950] require this convergence for all positive integer c.) The limit matrix of

$$n^{-1} \sum_{t=1}^{n} \underline{x}_t \underline{x}_t^T,$$

which is M_{00} in the notation of Koopmans et al., is required to be positive definite. Assuming that the error terms are independent, identically distributed with zero mean, positive variance σ^2 and finite moments of every order, and regarding the presample values $y_0, y_{-1}, \ldots, y_{1-G}$ as fixed, the least squares estimators of the parameters are shown to be consistent and asymptotically normally distributed with zero mean and covariance matrix equal to $\sigma^2 N^{-1}$, where N is as defined by term 2.3. If, in addition, $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are normally distributed, and if $\underline{b}$ is any other consistent estimator of the parameters, then the covariance matrix of the limiting distribution exceeds $\sigma^2 N^{-1}$ by a positive semidefinite matrix. The proof of this result is not given in Theil [1971].

Nocturne [1970] considers the model 2.1 where the exogenous

variables are no longer considered as constants, but as a stationary stochastic process independent of the error terms, such that

$$|\text{cov}\,(x^{i}_{t,h},\, x^{j}_{t+c,k})| \leq C\rho^{c} \qquad \begin{array}{l}\text{for } i,j=1,2,3,4,4+\delta;\\ \text{for } c=0,1,2,\ldots\,;\\ \text{for } h,k=1,\ldots,K;\end{array}$$

where $C \geq 0$, $0 \leq \rho < 1$, and $\delta > 0$. Stationarity implies the marginal distributions of $\underline{x}_t$ are identical for each t. If the marginal distribution is degenerate with mass one at e_t, then $e_t = e$, for all t. That is, if the exogenous variables are nonrandom, they must be constant. The error terms are assumed to be independently, normally distributed with expectation 0 and unknown variance σ^2. Nocturne also assumes that the process $\{y_t,\ t=1,2,\ldots\}$ is stationary, which implies the stability condition 2.2. Under these assumptions, the MLE of the parameters are shown to be consistent, asymptotically normally distributed, and efficient in the MP sense.

Other authors have dealt with various modifications of 2.1. Mann and Wald [1943] show, for the model 2.1 with no exogenous variables, that the MLE are consistent and asymptotically normally distributed, when the stability condition holds and the error variables are independently distributed with zero expectation and all other moments finite. Grenander and Rosenblatt [1957] also consider model 2.1 with no exogenous variables. Assuming that the stability condition holds, the error variables are independently, identically distributed with expectation zero and the next three moments finite, and the $\{y_t\}$ are second order stationary, the quasi-MLE (that is, the MLE computed under the assumption of normality of the errors) are shown to be consistent and asymptotically normally distributed.

Durbin [1960] derives a different criterion of optimality which the MLE of the parameters of 2.1 satisfy. If an estimator b of a scalar parameter β is defined by the linear equation

$$T_1 b + T_2 = 0 \tag{2.4}$$

where T_1 and T_2 are functions of the observations such that T_2/T_1 is independent of the unknown parameters, and

$$E(T_1\beta + T_2) = 0,$$

then 2.4 is called an unbiased linear estimating equation. An unbiased linear estimating equation $t_1 b + t_2 = 0$, with $E(t_1) = 1$ and

$$\text{var}(t_1\beta + t_2) \leq \text{var}(s_1\beta + s_2)$$

for all other unbiased linear estimating equations $s_1 b + s_2 = 0$, with $E(s_1) = 1$, is called a best unbiased linear estimating equation. Durbin derives a lower bound to $\text{var}(t_1\beta + t_2)$ similar to the derivation of the Cramér-Rao lower bound. He also shows that if the log likelihood function is a quadratic function of the unknown parameters, then the likelihood equations are best unbiased linear estimating equations. In this chapter we shall prove the optimality properties of Nocturne.

2.2 Optimality Properties for the MLE when the Exogenous Variables are Nonrandom

We will consider first the univariate model 2.1 where the exogenous variables are assumed to be nonrandom. The assumptions are similar to those of Koopmans, Rubin, and Leipnik [1950], but the results are stronger, since efficiency in the MP sense is stronger than efficiency in the sense of having the smallest asymptotic concentration ellipsoid.

Theorem 2.1

For model 2.1, assume the error terms are independent and normally distributed, with expectation zero and positive unknown variance σ^2. Assume that the stability condition 2.2 holds and that the exogenous variables are nonrandom, satisfying

1) $|x_{t,j}| \leq X$, a positive constant, for $j=1,\ldots,K$ and $t=1,2,\ldots$.

2) $n^{-1} \sum_{t=1}^{n} x_{t,h} x_{t+c,k}$ converges for $h,k=1,\ldots,K$ and for all c, uniformly in c.

Then the MLE of $\underline{\beta}$ and σ^2 are consistent, asymptotically normally distributed and efficient in the MP sense.

Proof:

The conditional density of y_j, given $y_{j-1},\ldots,y_1$ is normal with mean

$$\sum_{g=1}^{G} \beta_g y_{j-g} + \sum_{h=1}^{K} \beta_{G+h} x_{j,h}$$

and variance σ^2. Therefore the joint density of $y_1,y_2,\ldots,y_n$ is

$$L_n(\underline{\beta},\sigma^2) = \prod_{t=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\{-(2\sigma^2)^{-1}(y_t - \sum_{g=1}^{G} \beta_g y_{t-g} - \sum_{h=1}^{K} \beta_{G+h} x_{t,h})^2\}$$

The second derivatives of the log likelihood function, with respect to the parameters to be estimated, are:

$$2.5 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{\partial\beta_i\ \partial\beta_j} = (n\ \sigma^2)^{-1} \sum_{t=1}^{n} y_{t-i} y_{t-j} \qquad i,j=1,\ldots,G$$

$$2.6 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{\partial\beta_{G+i}\ \partial\beta_{G+j}} = (n\sigma^2)^{-1} \sum_{t=1}^{n} x_{t,i} x_{t,j} \qquad i,j=1,\ldots,K$$

$$2.7 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{(\partial\sigma^2)^2} = (-2\sigma^4)^{-1} + (n\sigma^6)^{-1} \sum_{t=1}^{n} \varepsilon_t^2$$

$$2.8 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{\partial\beta_i\ \partial\beta_{G+j}} = (n\ \sigma^2)^{-1} \sum_{t=1}^{n} x_{t,j} y_{t-i} \qquad i=1,\ldots G;j=1,\ldots,K$$

$$2.9 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{\partial\sigma^2\ \partial\beta_i} = (n\sigma^4)^{-1} \sum_{t=1}^{n} \varepsilon_t y_{t-i} \qquad i=1,\ldots G$$

$$2.10 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\beta},\sigma^2)}{\partial\sigma^2\ \partial\beta_{G+j}} = (n\sigma^4)^{-1} \sum_{t=1}^{n} \varepsilon_t x_{t,j} \qquad j=1,\ldots,K$$

It is necessary to show that the expected value of 2.5 through 2.10 converge to functions of the parameters which are continuous, and that the variances are bounded above by terms which go to 0 as $n \to \infty$, uniformly for values of the unknown parameters in any compact set.

First we will undertake the proofs for term 2.5. The variables

$$\Phi_s(t) = \beta_1\Phi_s(t-1) + \beta_2\Phi_s(t-2) + \ldots + \beta_{t-s}\Phi_s(s)$$

$$\Phi_t(t) = 1, \quad \Phi_t(t-h) = 0 \quad \text{for all} \quad h > 0.$$

Using the last restriction we may write the above system as

$$\Phi_s(t) = \beta_1\Phi_s(t-1) + \beta_2\Phi_s(t-2) + \ldots\ldots + \beta_G\Phi_s(t-G)$$

for $0 \leq s < t$, where the Φ's satisfy, for $s=1,\ldots,t$,

$$\Phi_s(s)=1,\ \Phi_s(s-1)=0,\ \ldots\ldots\ldots\ldots,\Phi_s(s-G+1)=0$$

$$\Phi_0(0)=y_0,\Phi_0(-1)=0,\ \ldots\ldots\ldots\ldots,\Phi_0(-G+1)=y_{-G+1}.$$

The solution to this system of difference equations is

$$\Phi_s(t) = P_{1s}(t)\rho_1^t + \ldots\ldots +P_{\ell s}(t)\rho_\ell^t \qquad \text{for } 0\leq s<t$$

where $\rho_1,\ldots,\rho_\ell$ are the distinct roots of 2.2 and $P_{is}(t)$ is a polynomial in t whose degree is one less than the multiplicity of the root ρ_i. These polynomials are determined uniquely from the initial conditions and the multiplicity of the roots. Evaluating the polynomials at the initial conditions, we have

$$1 = \Phi_s(s) = \sum_{i=1}^{\ell} P_{is}(s)\rho_i^s$$

$$0 = \Phi_s(s-j) = \sum_{i=1}^{\ell} P_{is}(s-j)\rho_i^{s-j} \qquad j=1,\ldots,G-1$$

Thus $P^*_{is}(t) = P_{i1}(t-s+1)/\rho_i^{s-1}$ is a polynomial of the same degree as $P_{is}(t)$ and also satisfies the initial conditions. Therefore we must have $P_{is}(t) = P^*_{is}(t)$ and

$$\Phi_s(t) = \sum_{i=1}^{\ell} P_{i1}(t-s+1)\rho_i^{t-s+1} \qquad \text{for } 0 < s < t. \tag{2.12}$$

In the computations which follows, we will assume without loss of generality that all G roots are distinct, and therefore $P_{i1}(t-s+1)$ has degree zero and is not a function of $t-s$. Therefore let $\lambda_i = P_{i1}(t-s+1)$ for $i=1,\ldots G$. The parameters λ_i and ρ_i are continuous as functions of $\beta_1,\beta_2,\ldots,\beta_G$, $y_0,y_{-1},\ldots,y_{-G+1}$. When the dependence on $\underline{\beta}$ is to be emphasized, the parameters will be written as $\lambda_i(\underline{\beta})$ and $\rho_i(\underline{\beta})$.

The proof that the expectation of 2.5 converges follows. Assuming that $i \geq j$,

$$E\left(n^{-1}\sum_{t=1}^{n} y_{t-i}y_{t-j}\right) = n^{-1}\sum_{t=1}^{j} y_{t-i}y_{t-j} + n^{-1}\sum_{t=j+1}^{i} y_{t-i}E(y_{t-j}) + n^{-1}\sum_{t=i+1}^{n}\sum_{k=0}^{t-i}\sum_{m=0}^{t-j}\Phi_k(t-i)\Phi_m(t-j)\,E(\phi_k\phi_m) \tag{2.13}$$

The first term is not a function of $\underline{\beta}$ and goes to zero as $n \to \infty$. The summand of the second term of 2.13 is bounded above by a continuous function of $\underline{\beta}$, and therefore the second term goes to zero uniformly for $(\underline{\beta},\sigma^2)$ in a compact set. To evaluate the third term, we calculate

$$2.14 \qquad E(\phi_k\phi_m) = \left(\sum_{a=1}^{K}\beta_{G+a}x_{k,a}\right)\left(\sum_{b=1}^{K}\beta_{G+b}x_{m,b}\right) + \delta_{km}\sigma^2 \qquad k,m>0$$

$$= \sum_{a=1}^{K}\beta_{G+a}x_{k,a} \qquad k>0, m=0$$

$$= 1 \qquad k=m=0.$$

So the third term of 2.13 equals

$$2.15 \qquad n^{-1}\sum_{t=i+1}^{n}\Phi_0(t-i)\Phi_0(t-j) + n^{-1}\sum_{t=i+1}^{n}\Phi_0(t-i)\sum_{m=1}^{t-j}\Phi_m(t-j)\left(\sum_{a=1}^{K}\beta_{G+a}x_{m,a}\right)$$

$$+ n^{-1}\sum_{t=i+1}^{n}\Phi_0(t-j)\sum_{k=1}^{t-i}\Phi_k(t-i)\left(\sum_{a=1}^{K}\beta_{G+a}x_{k,a}\right)$$

$$+ n^{-1}\sum_{t=i+1}^{n}\sum_{k=1}^{t-i}\sum_{m=1}^{t-j}\Phi_k(t-i)\Phi_m(t-j)\left[\left(\sum_{a=1}^{K}\beta_{G+a}x_{k,a}\right)\left(\sum_{b=1}^{K}\beta_{G+b}x_{m,b}\right) + \delta_{km}\sigma^2\right]$$

where δ_{km} is 1 if $k=m$ and is otherwise 0. The convergence, uniformly in $\underline{\beta}$, of the second and third terms of 2.15 may be proven in the same manner as the following proof for the convergence to zero of the first term.

$$2.16 \qquad n^{-1}\sum_{t=i+1}^{n}\Phi_0(t-i)\Phi_0(t-j) = \sum_{c=1}^{G}\sum_{d=1}^{G}\lambda_c\,\lambda_d\left(n^{-1}\sum_{t=i+1}^{n}\rho_c^{\,t-i+1}\,\rho_d^{\,t-j+1}\right)$$

$$= n^{-1}\left(\sum_{c=1}^{G}\sum_{d=1}^{G}\lambda_c\,\lambda_d\,\rho_d^{\,i-j}[(\rho_c\rho_d)^2 - (\rho_c\,\rho_d)^{n-i+2}]/(1-\rho_c\,\rho_d)\right)$$

which converges uniformly to 0 for $(\underline{\beta},\sigma^2)$ in a compact set, since ρ_k is less than one in absolute value, for $k=1,\ldots,G$, and since

λ_k and ρ_k are continuous functions of $\underline{\beta}$, for $k=k,\ldots,G$.

It remains to show that the last term of 2.15 converges uniformly in $\underline{\beta}$. That term may be written as

$$2.17 \quad \sum_{c=1}^{G} \sum_{d=1}^{G} \lambda_c \lambda_d \sum_{a=1}^{K} \sum_{b=1}^{K} \beta_{G+a} \beta_{G+b} *$$

$$\left\{ n^{-1} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \sum_{m=1}^{t-j} \rho_c^{t-i-k+1} \rho_d^{t-j-m+1} x_{k,a} x_{m,b} \right\}$$

$$+ \sum_{c=1}^{G} \sum_{d=1}^{G} \lambda_c \lambda_d (\sigma^2 n^{-1} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \rho_c^{t-i-k+1} \rho_d^{t-j-k+1})$$

The last term of 2.17 converges uniformly, for $\underline{\beta}$ in a compact set, to

$$\sigma^2 \sum_{c=1}^{G} \sum_{d=1}^{G} \lambda_c \lambda_d \rho_d^{i-j+1} \rho_c / (1-\rho_c \rho_d).$$

Assumption 2 implies that the first term of 2.17 converges uniformly. The factor of this term in brackets may be written as

$$2.18 \quad \sum_{k=1}^{n-i} \sum_{m=1}^{k+r-1} \rho_c^{k} \rho_d^{k+r-m} (n^{-1} \sum_{q=0}^{n-i-k} x_{q+1,a} x_{q+m+1,b})$$

$$+ \sum_{k=1}^{n-1} \sum_{m=0}^{n-i-k} \rho_c^{k} \rho_d^{m+k+r} (n^{-1} \sum_{q=0}^{n-i-m-k} x_{q+m+1,b} x_{q+1,a})$$

We will show that the first term is a Cauchy sequence in n. Denote by $S(n-i,m)$ the term

$$n^{-1} \sum_{q=0}^{n-i} x_{q+1,a} x_{q+m+1,b}$$

Choose n,p sufficiently large so that, for all m, $|S(n-i,m)-S(p-i,m)|<\varepsilon$. Denoting the first term of 2.18 by $T(n)$, we have

$$|T(n)-T(p)| \leq \sum_{k=1}^{p-i} \sum_{m=1}^{k+r-1} |\rho_c|^k \, |\rho_d|^{k+r-m}(\varepsilon + X^2k/n + X^2k/p)$$

$$+ X^2 \sum_{k=p-i+1}^{n-i} \sum_{m=1}^{k+r-1} |\rho_c|^k \, |\rho_d|^{k+r-m}$$

where X is defined in assumption 1. It is clear that this upper bound goes to zero as $n,p \to \infty$. Moreover, the convergence of the first term of 2.18 is uniform for $\underline{\beta}$ in a compact set, since ρ_c, ρ_d are continuous functions of $\underline{\beta}$. A similar argument shows that the second term of 2.18 converges uniformly for $\underline{\beta}$ in a compact set.

We will now show that the variance of 2.5 goes to zero, as $n \to \infty$, uniformly for $\underline{\beta},\sigma^2$ in a compact set. For $i \geq j$,

$$(n^2\sigma^4)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \operatorname{cov}(y_{t-i}y_{t-j},\ y_{s-i}y_{s-j})$$

2.19 $$= (n^2\sigma^4)^{-1} \sum_{t=j+1}^{i} \sum_{s=j+1}^{i} y_{t-i}y_{s-i} \operatorname{cov}(y_{t-j},\ y_{s-j})$$

$$+ 2(n^2\sigma^4)^{-1} \sum_{t=j+1}^{i} \sum_{s=i+1}^{n} y_{t-i} \operatorname{cov}(y_{t-j},\ y_{s-i}y_{s-j})$$

$$+ (n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \operatorname{cov}(y_{t-i}y_{t-j},\ y_{s-i}y_{s-j}).$$

The first two terms go to zero as required since each term of the double summand is bounded in absolute value by a constant, for $\underline{\beta},\sigma^2$

in a compact set. The uniform convergence to zero of the last term remains to be shown. The last term may be written as

$$2.20 \quad (n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \sum_{p=0}^{t-i} \sum_{q=0}^{t-j} \sum_{r=0}^{s-i} \sum_{v=0}^{s-j} \phi_p(t-i)\phi_q(t-j)\phi_r(s-i)\phi_v(s-j) *$$

$$\{E(\phi_p\phi_q\phi_r\phi_v) - E(\phi_p\phi_q)E(\phi_r\phi_v)\}$$

We can evaluate the term in brackets in 2.20 as

$E(\phi_p\phi_q\phi_r\phi_v) - E(\phi_p\phi_q)E(\phi_r\phi_v)$

$= \sigma^4 + \sigma^2(m_p^2 + m_q^2)$	if $p=r$, $q=v$, $p\neq q$ or if $p=v$, $q=r$, $p\neq q$
$= \sigma^2 m_q m_v$	if $p=r$, $q\neq v$, $q\neq p\neq v$
$= \sigma^2 m_q m_r$	if $p=v$, $q\neq r$, $q\neq p\neq r$
$= \sigma^2 m_p m_r$	if $q=v$, $p\neq r$, $p\neq q\neq r$
$= \sigma^2 m_p m_v$	if $q=r$, $p\neq v$, $p\neq q\neq v$
$=2\sigma^2 m_r m_p$	if $p=q=v$, $r\neq p$
$=2\sigma^2 m_v m_p$	if $p=q=r$, $v\neq p$
$=2\sigma^2 m_p m_q$	if $q=r=v$, $p\neq q$
$=2\sigma^2 m_p m_q$	if $p=r=v$, $q\neq p$
$=2\sigma^4+4\sigma^2 m_p^2$	if $p=q=r=v$
$= 0$	otherwise.

where $E(\phi_p) = m_p = \sum_{a=1}^{K} \beta_{G+a} x_{p,a}$ for $p \geq 1$. The terms involving $\phi_0 = 1$ may be ignored, since they are bounded above by a continuous function of $(\underline{\beta},\sigma^2)$ which goes to zero as $n \to \infty$. We will prove the same is true for each of the above terms, assuming without loss of

generality that $t \geq s$ and $j \geq i$. Let A,B,C,D represent distinct elements of $\{t-i,t-j,s-i,s-j\}$. A and B are said to be of the same type if A,B are in the set $\{t-i,t-j\}$ or the set $\{s-i,s-j\}$. The sums involved are either single, double, or triple sums over p,q,r and v, since $\mathrm{cov}(\phi_p\phi_q,\phi_r\phi_v) = 0$ unless at least two indices are equal. Let f(S,X) be an upper bound for each of the above terms, when $(\underline{\beta},\sigma^2)$ are in the compact set S. X was defined in assumption 1) as an upper bound for the exogenous variables.

We will note first that

$$|\Phi_p(t-i)| \leq [\sum_{c=1}^{G} |\lambda_c|]\rho^{t-i-p+1}$$

where $\rho = \max\{\rho_1,\rho_2,\ldots,\rho_G\}$, and ρ is a continuous function of $\underline{\beta}$. The single summand is of the form

$$\sum_{p=1}^{s-j}\Phi_p(t-i)\Phi_p(t-j)\Phi_p(s-i)\Phi_p(s-j)\ f_p(\sigma^2,\underline{\beta},\underline{x})$$

where f_p is a continuous function of its arguments. This summation is bounded above in absolute value by

$$f(S,X)\ (\sum_{c=1}^{G} |\lambda_c|)^4 \rho^{2(t-s)+2(j-i)}(\rho^4-\rho^{4(s-j+2)})(1-\rho^4)^{-1}$$

which is bounded above by a continuous function of the parameters and exogenous variables times ρ^{t-s}.

The form of the double summand is

$$\sum_{p=1}^{A}\ \sum_{q=1}^{B}\Phi_p(A)\Phi_q(B)\Phi_r(C)\Phi_v(D)\ f_{pq}(\sigma^2,\underline{\beta},\underline{x})$$

where f_{pq} is a covariance term. Hence it is bounded above by $f(S,X)$. Assume first that A,B are of the same type. Then C,D are of the same type. Consider first the case where $r=v=p$, which will be argued identically to the case $r=v=q$, by symmetry. In this case the summand equals

$$2.21 \qquad \sum_{p=1}^{A} \sum_{q=1}^{B} \phi_p(A)\phi_q(B)\phi_q(C)\phi_q(D)\ f_{pq}(\sigma^2,\underline{\beta},\underline{x})$$

$$\leq f(S,X)\ (\sum_{c=1}^{G} |\lambda_c|)^4\ \rho^4(1-\rho)^{-1}\ (1-\rho^3)^{-1}\ \rho^{C+D-2B}$$

C and D are both of the type opposite from B, and $C > B$, $D > B$. Since $t \geq s$ and $j \geq i$, either $t-s \geq j-i$, which implies $t-i \geq t-j \geq s-i \geq s-j$, or $t-s \leq j-i$, which implies $t-i \geq s-i \geq t-j \geq s-j$. Therefore we must have C,D in $\{t-i,t-j\}$ and B in $\{s-i,s-j\}$, so that $C+D-2B = 2(t-s)-(j-i)$ if $B=s-i$ and $C+D-2B = 2(t-s)+(j-i)$ if $B=s-j$. So an upper bound for the summand is

$$f(S,X)(\sum_{c=1}^{G} |\lambda_c|)^4\rho^4(1-\rho)^{-1}(1-\rho^3)^{-1}\ \rho^{t-s}$$

Now, with A,B of the same type, suppose r and t are unequal, and, without loss of generality, $p=r$ and $q=t$, which implies $C \geq A$ and $D \geq B$. Then an upper bound for the absolute value of the summand is

$$f(S,X)(\sum_{c=1}^{G}|\lambda_c|)^4\rho^4(1-\rho^2)^{-2}\ \rho^{C-A+D-B}$$

The restrictions $C \geq A$, $D \geq B$, A and B of the same type imply that

A,B are in $\{s-i,s-j\}$ and C,D are in $\{t-i,t-j\}$, which implies that

$$\rho^{C-A+D-B} \leqq \rho^{2(t-s)} \leqq \rho^{(t-s)}$$

So an upper bound for the absolute value of this summand is

$$f(S,X)\left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^4(1-\rho^2)^{-2} \rho^{t-s}$$

Now the double summand with A,B of different types will be considered. Suppose without loss of generality that C is of the same type as A and D is of the same type as B. If $r=p$ and $v=q$, then $\mathrm{cov}(\phi_p\phi_p, \phi_q\phi_q) = 0$, for $p\neq q$, so this term does not appear among the covariance terms. Now suppose that $r=q$ and $v=p$, which implies $C > B$ and $D > A$. Then

$$\left|\sum_{p=1}^{A} \sum_{q=1}^{B} \Phi_p(A)\Phi_p(D)\Phi_q(D)\Phi_q(C)\ f_{pq}(\sigma^2,\underline{\beta},\underline{x})\right)$$

$$\leqq f(S,X)\left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^4(1-\rho^2)^{-2} \rho^{C-B+D-A}$$

This term appears only when $t-i\geqq s-i\geqq t-j\geqq s-j$, or $j-i\geqq t-s$, and in this case $C-B+D-A = 2(j-i) \geqq t-s$. Therefore an upper bound for the absolute value of the summand is

$$f(S,X)\left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^2(1-\rho^2)^{-1} \rho^{t-s}$$

Retaining the assumption that A,B are of different types, suppose $p=r=v$, and $C\geqq A$, $C\geqq D$. Then

$$\sum_{p=1}^{A}\sum_{q=1}^{B}\Phi_p(A)\phi_p(C)\phi_p(D)\Phi_q(B)\ f_{pq}(\sigma^2,\underline{\beta},\underline{x})$$

$$\leq f(S,X)\ (\sum_{c=1}^{G}|\lambda_c|)^4\ \rho^4(1-\rho)^{-1}(1-\rho^3)^{-1}\ \rho^{D+C-2A}$$

If $t-i\geq t-j\geq s-i\geq s-j$ (i.e., if $t-s\geq j-i\geq 0$) then the restriction that $C\geq A$ and $D\geq A$, C and A of the same type, implies $C=s-i$ and $A=s-j$. If $D=t-i$ then $D+C-2A = t-s + 2(j-i) \geq t-s$. If $D=t-j$ then $D+C-2A = t-s +j-i\geq t-s$. If $t-i>s-i>t-j>s-j$ (i.e., if $0<t-s<j-i$) then there are several possibilities. First let $C=s-i$ and $A=s-j$. If $D=t-i$, then $D+C-2A = t-s+2(j-i) \geq t-s$. If $D=t-j$, then $D+C-2A = t-s + j-i\geq t-s$. Secondly, let $C=t-i$ and $A=t-j$. This implies that $D=s-i$ and $B=s-j$, so that $D+C-2A = j-i + s-i - (t-j) \geq j-i \geq t-s$. Having exhausted all possibilities we may conclude that an upper bound is

$$f(S,X)(\sum_{c=1}^{G}|\lambda_c|)^4\ \rho^4(1-\rho)^{-1}(1-\rho^3)^{-1}\rho^{t-s}$$

This ends the analysis of terms which are double summands.

Now let us consider terms which are triple summands. These terms are of the form

$$\sum_{p=1}^{A}\sum_{q=1}^{B}\sum_{r=1}^{C}\Phi_p(A)\Phi_q(B)\Phi_r(C)\Phi_p(D)\ f_{pqr}(\sigma^2,\underline{\beta},\underline{x})$$

If A,D are of the same type, $f_{pqr}(\sigma^2,\underline{\beta},\underline{x}) = \mathrm{cov}(\phi_p^2,\phi_r\phi_q) = 0$ when $p\neq q\neq r$. If the summand is to be nonzero, D must be of a different type than A. In this case an upper bound of the absolute value of the summand is

$$f(S,X)\left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^4(1-\rho^2)^{-1}(1-\rho)^{-2} \rho^{t-s}$$

This completes the determination of upper bounds for $\operatorname{cov}(y_{t-i}y_{t-j}, y_{s-i}y_{s-j})$ with $t \geqq s$ and $j \geqq i$. We may conclude, by applying Lemma 1.1, that, for $\underline{\beta},\sigma^2$ in a compact set S,

$$\left|n^{-2} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \operatorname{cov}(y_{t-i}\, y_{t-j},\, y_{s-i}\, y_{s-j})\right|$$

$$\leqq \text{const.}\left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^4(1-\rho^2)^{-1}(1-\rho)^{-2}\left(n^{-2} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \rho^{|t-s|}\right)$$

which goes to zero as $n \to \infty$, uniformly for $\underline{\beta},\sigma^2$ in a compact set S.

Next we will show that the expected value of 2.6 converges as $n \to \infty$ uniformly for $\underline{\beta},\sigma^2$ in a compact set. Assumption 2) implies this. Since the $\{x_{t,j}\}$ are nonrandom, the variance of 2.6 is zero.

For term 2.7, we determine that its expected value is $(2\sigma^4)^{-1}$, which is not a function of n, but is clearly continuous in σ^2. The variance of 2.7 is $3(n\sigma^8)^{-1}$, which goes to zero uniformly for σ^2 in a compact set.

For term 2.8, we calculate the expected value as

$$2.22 \quad (n\sigma^2)^{-1} \sum_{t=1}^{i} x_{t,j} y_{t-i} + (n\sigma^2)^{-1} \sum_{t=i+1}^{n} x_{t,j}\left(\Phi_0(t-i) + \sum_{k=1}^{t-i} \Phi_k(t-i) m_k\right)$$

where, for $k \geq 1$, $m_k = \sum_{a=1}^{K} \beta_{G+a} x_{k,a} = E(\phi_k)$. Since the first term of 2.22 is a sum of constants times n^{-1}, it goes to zero uniformly in $\underline{\beta},\sigma^2$. The second term may be written as

$$2.23 \qquad \sigma^{-2} \sum_{c=1}^{G} \sum_{a=1}^{K} \lambda_c \beta_{G+a} \Big(\sum_{k=1}^{n-i} \rho_c^k \, [n^{-1} \sum_{t=k}^{n-i} x_{t+i,j} x_{t-k+1,a}] \Big)$$

$$+ \sum_{c=1}^{G} \lambda_c \, [(n\sigma^2)^{-1} \sum_{t=i+1}^{n} x_{t,j} \rho_c^{t-i+1}]$$

By assumption 1, the second term of 2.23 goes to zero uniformly for $\underline{\beta}, \sigma^2$ in a compact set. Using assumption 2) and arguments similar to those used for term 2.18, it is possible to show that the first term of 2.23 converges, as $n \to \infty$, uniformly for $\underline{\beta}$ in a compact set. Therefore we conclude that the expected value of 2.8 converges in a similar manner.

We must also show that the variance of 2.8 converges uniformly to zero. The variance of 2.8 is

$$2.24 \qquad (n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} x_{t,j} \, x_{s,j} \operatorname{cov}(y_{t-i}, y_{s-i})$$

For $t \geq s$, $\operatorname{cov}(y_{t-i}, y_{s-i}) = \sigma^2 \sum_{k=1}^{s-i} \Phi_k(t-i)\Phi_k(s-i)$

$$\leq \sigma^2 \Big(\sum_{c=1}^{G} |\lambda_c| \Big)^4 \rho^2 (1-\rho^2)^{-1} \rho^{t-s}$$

Therefore, by Lemma 1.1, term 2.24 goes to zero, uniformly for $\underline{\beta}, \sigma^2$ in a compact set.

To compute the expected value of 2.9, note that y_{t-i} is a sum of error terms ε_k for $k \leq t-i$. Since the error terms are independent, the expected value of 2.9 is zero. The variance of 2.9 is

$$2.25 \quad (n^2\sigma^8)^{-1} \left\{ \sum_{t=1}^{i} y_{t-i}^2 \sigma^2 + \sum_{t=i+1}^{n} \Phi_0(t-i)^2\sigma^2 \right.$$

$$+ \sum_{t=i+1}^{n} 2\Phi_0(t-i) \sum_{p=1}^{t-i} \Phi_p(t-i)\sigma^2 m_p$$

$$\left. + \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \sum_{p=1}^{t-i} \Phi_k(t-i)\Phi_p(t-i)[\delta_{kp}\sigma^4 + m_k m_p \sigma^2] \right\}$$

which goes to zero uniformly for $\underline{\beta},\sigma^2$ in a compact set since 2.25 is bounded above by a continuous function of $\underline{\beta},\sigma^2$ times n^{-1}.

It is clear that the expected value of 2.10 is zero and the variance of 2.10 is

$$(n\sigma^6)^{-1} (n^{-1} \sum_{t=1}^{n} x_{t,j}^2)$$

Since the exogenous variables are bounded, the variance goes to zero as $n \to \infty$, uniformly for σ^2 in a compact set.

Since the assumptions for theorems 1.1 and 1.2 are satisfied, theorem 1.3 may be applied to conclude that the MLE of $(\underline{\beta},\sigma^2)$ are consistent, asymptotically normally distributed, and asymptotically efficient in the MP sense. The asymptotic covariance matrix is the matrix of the stochastic limits of 2.5 through 2.10.

2.3 Optimality Properties for the MLE when the Exogenous Variables are Random

Asymptotic properties for the MLE for models with lagged dependent variables may also be derived when the exogenous variables are assumed to be random. Nocturne's assumption that $\{y_t, t=1,2,\ldots\}$ is a stationary

process and that the covariances of the third and fourth powers of the exogenous variables are decreasing exponentially with their difference in time, are not necessary. Our assumptions regarding the exogenous variables involve powers no higher than two.

Theorem 2.2

For the model 2.1, assume the exogenous variables $\{x_t, t=1,2,\ldots\}$ are identically distributed, independently of the error sequence $\{\varepsilon_t, t=1,2,\ldots\}$, which is itself normally distributed with mean 0 and variance σ^2, independently. Assume that the stability condition 2.2 holds. Assume also that

1) $|\mathrm{cov}(x_{t,k}, x_{s,m})| = |R_{km}(|t-s|)| \leq C\eta^{|t-s|}$

   for $k,m=1,\ldots,K;\ t,s=1,2,\ldots$

2) $|\mathrm{cov}(x_{t,k}, x_{s,m}x_{w,i})| \leq C\eta^{M}$, where $M=\max\{|t-s|,|t-w|\}$

   for $k,m,i=1,\ldots,K;\ t,s,w=1,2,\ldots$

3) $|\mathrm{cov}(x_{t,k}x_{v,j}, x_{s,m}x_{w,i})| \leq C\eta^{M}$, where

   $M=\max\{|t-s|,|t-w|,|v-s|,|v-w|\}$, for $k,m,i,j=1,\ldots,K;$

   for $t,s,w,v=1,2,\ldots$

where C is a positive constant and $0 \leq \eta < 1$. Then the MLE of $\underline{\beta},\sigma^2$ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Denote the moments of the process $\{x_{t,k}\}$ as

$$E(x_{t,k}) = \mu_k$$

$$\mathrm{cov}(x_{t,h}, x_{s,k}) = R_{hk}(|t-s|) \quad \text{for} \quad h,k=1,\ldots,K, \text{ for } t,s=1,2,.$$

Since the process $\{\underset{\sim}{x}_t\}$ is independent of the error terms, and its density is not a function of the parameters to be estimated, the second derivatives of the log likelihood function are identical with 2.5 through 2.10. As in the proof of theorem 2.1, y_t can be expressed as

$$y_t = \sum_{k=0}^{t} \Phi_k(t)\phi_k$$

where $\phi_0 = 1$

and $$\phi_k = \sum_{a=1}^{K} \beta_{G+a} x_{k,a}$$

and $$\Phi_k(t) = \sum_{m=1}^{G} \lambda_m \rho_m^{t-k+1} \quad \text{for} \quad k=0,\ldots,t;\ t=1,\ldots,n$$

Analyzing term 2.5 first, we calculate, that for $t \geq i \geq j$,

2.26 $$E(y_{t-i}y_{t-j}) = \Phi_0(t-i)\,\Phi_0(t-j)$$

$$+ \left(\sum_{a=1}^{K} \beta_{G+a}\mu_a\right)\left\{\Phi_0(t-i)\sum_{m=1}^{t-j}\Phi_m(t-j) + \Phi_0(t-j)\sum_{k=1}^{t-i}\Phi_k(t-i)\right\}$$

$$+ \sigma^2 \sum_{k=0}^{t-i} \Phi_k(t-i)\,\Phi_k(t-j)$$

$$+ \sum_{a=1}^{K}\sum_{b=1}^{K} \beta_{G+a}\,\beta_{G+b}\left\{\sum_{k=1}^{t-i}\sum_{m=1}^{t-j}\Phi_k(t-i)\Phi_m(t-j)[R_{ab}(|k-m|) + \mu_a\mu_b\right\}$$

Summing 2.26 over t from $i+1$ to n, and dividing by n, yields the

term whose convergence we must show. Clearly the first two terms go to zero as $n \to \infty$, uniformly for $\underline{\beta},\sigma^2$ in a compact set. The average over t from $i+1$ to n of the third term of 2.26 is identical to the second term of 2.17. Lemmas 1.3 and 1.4 imply that the fourth term converges as $n \to \infty$ uniformly for $\underline{\beta},\sigma^2$ in a compact set.

To show that the variance of 2.5 converges to 0 uniformly in $\underline{\beta},\sigma^2$ we have only to check the convergence of 2.20, under the assumption that the sequence of exogenous variables is random. With m_p defined as in the proof of theorem 2.1,

$$m_p = \sum_{a=1}^{K} \beta_{G+a} x_{p,a},$$

the covariance function required may be expressed as

$$\begin{aligned} 2.27 \qquad \operatorname{cov}(\phi_p\phi_q,\phi_r\phi_v) &= \operatorname{cov}(m_p m_q, m_r m_v) + \operatorname{cov}(\varepsilon_p\varepsilon_q,\varepsilon_r\varepsilon_v) \\ &+ E(m_p m_r)\, E(\varepsilon_q\varepsilon_v) + E(m_p m_v)\, E(\varepsilon_q\varepsilon_r) \\ &+ E(m_q m_r)\, E(\varepsilon_p\varepsilon_v) + E(m_q m_v)\, E(\varepsilon_p\varepsilon_r). \end{aligned}$$

Convergence for each of the terms of 2.27 will be verified. The first term can be written as

$$\operatorname{cov}(m_p m_q, m_r m_v) = \sum_{a=1}^{K}\sum_{b=1}^{K}\sum_{c=1}^{K}\sum_{d=1}^{K} \beta_{G+a}\,\beta_{G+b}\,\beta_{G+c}\,\beta_{G+d} \;*$$

$$\operatorname{cov}(x_{p,a}\, x_{q,b},\, x_{r,c}\, x_{v,d})$$

whose absolute value is, by assumption 3, bounded above by

$$\left(\sum_{a=1}^{K} |\beta_{G+a}|\right)^4 C\eta^{|p-r|}$$

since $\max\{|p-r|,|p-v|,|q-r|,|q-v|\} \geq |p-r|$, and $0 \leq \eta < 1$. So, for the first term of 2.27, we have

$$\left|(n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \sum_{p=1}^{t-i} \sum_{q=1}^{t-j} \sum_{r=1}^{s-i} \sum_{v=1}^{s-j} \Phi_p(t-i)\, \Phi_q(t-j)\, \Phi_r(s-i)\, \Phi_v(s-j) * \right.$$
$$\operatorname{cov}(m_p\, m_q,\ m_r\, m_v)$$

$$\leq C\left(\sum_{a=1}^{K} |\beta_{G+a}|\right)^4 \left(\sum_{c=1}^{G} |\lambda_c|\right)^4 \rho^2 (1-\rho)^{-2} (n^2\sigma^4)^{-1} \pi^2 (1-\pi^2)^{-1} *$$
$$\left\{ \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \left(|t-s|\pi^{|t-s|} + (3-\pi^2)(1-\pi^2)^{-1}\pi^{|t-s|}\right)\right\}$$

where $\pi = \max\{|\rho_1|,|\rho_2|,\ldots,|\rho_G|,\eta\}$. By lemmas 1.1, 1.2 this upper bound converges to zero uniformly for $\underline{\beta},\sigma^2$ in a compact set.

For the second term of 2.27, we note that

$$\begin{aligned}
\operatorname{cov}(\varepsilon_p\varepsilon_q,\ \varepsilon_r\varepsilon_v) &= 2\sigma^4 && \text{if } p=q=r=v \\
&= \sigma^4 && \text{if } p=r, q=v, p\neq q \text{ or } p=v, q=r, p\neq q \\
&= 0 && \text{otherwise}
\end{aligned}$$

Therefore, we have that, if $i \geq j$

$$(n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \sum_{p=1}^{t-i} \sum_{q=1}^{t-j} \sum_{r=1}^{s-i} \sum_{v=1}^{s-j} \Phi_p(t-i)\, \Phi_q(t-j)\, \Phi_r(s-i)\, \Phi_v(s-j) *$$
$$\operatorname{cov}(\varepsilon_p\varepsilon_q,\ \varepsilon_r\varepsilon_v)$$

$$\leq 4\left(\sum_{m=1}^{G} |\lambda_m|\right)^4 \rho^4 (1-\rho^2)^{-2} \left(n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} \rho^{|t-s|}\right)$$

which, by lemma 1.1, goes to zero uniformly for $\underline{\beta},\sigma^2$ in a compact set. Thus the mean and variance of 2.5 converge as required.

The expected value of term 2.6 is

$$\sigma^{-2}(R_{ij}(0) + \mu_i\mu_j)$$

which is not a function of n. Also, the variance of 2.6 is

$$(n^2\sigma^4)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathrm{cov}(x_{t,i}\, x_{t,j},\ x_{s,i}\, x_{s,j})$$

which is bounded above in absolute value by

$$C(n^2\sigma^4)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} \eta^{|t-s|}$$

which goes to zero as $n \to \infty$, uniformly for $\underline{\beta},\sigma^2$ in a compact set, by lemma 1.1.

The mean and variance of 2.7 converge as indicated in the proof of theorem 2.1.

The expectation of 2.8 is

$$(n\sigma^2)^{-1} \sum_{t=1}^{i} \mu_j y_{t-i} + (n\sigma^2)^{-1} \sum_{t=i+1}^{n} \Phi_0(t-i)\mu_j \tag{2.28}$$

$$+ \sum_{m=1}^{G} \sum_{a=1}^{K} \lambda_m\, \beta_{G+a}\, (n\sigma^2)^{-1} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \rho_m^{t-i-k+1}[\mu_j\mu_a + R_{ja}(|t-k|)]$$

The first and second terms of 2.28 go to zero uniformly for $\underline{\beta},\sigma^2$ in a compact set. Lemma 1.3 implies the third term of 2.28 converges as

$n \to \infty$. The continuity of λ_m and ρ_m as functions of $\underline{\beta}$ implies uniform convergence of this term.

It remains to show that the variance of 2.8 goes to zero uniformly in $\underline{\beta},\sigma^2$. We can write the variance as

$$(n^2\sigma^4)^{-1} \sum_{t=1}^{i} \sum_{s=1}^{i} y_{t-i}\, y_{s-i}\, \mathrm{cov}(x_{t,j},\, x_{s,j}) \tag{2.29}$$

$$+\ 2(n^2\sigma^4)^{-1} \sum_{t=1}^{i} \sum_{s=i+1}^{n} y_{t-i}\, \mathrm{cov}(x_{t,j},\, x_{s,j}\, y_{s-i})$$

$$+\ (n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{s=i+1}^{n} \mathrm{cov}(x_{t,j}\, y_{t-i},\, x_{s,j}\, y_{s-i})$$

The first term of 2.29 clearly goes to zero as required. The second term goes to zero uniformly if

$$\mathrm{cov}(x_{t,j},\, x_{s,j} y_{s-i}) = \sum_{k=0}^{s-i} \phi_k(s-i) \sum_{a=1}^{K} \beta_{G+a}\, \mathrm{cov}(x_{t,j},\, x_{s,j} x_{s,a})$$

is bounded uniformly in s by a continuous function of $\underline{\beta},\sigma^2$. By assumption 2, this is true. For the third term of 2.29, we can write

$$\mathrm{cov}(x_{t,j} y_{t-i},\, x_{s,j} y_{s-i}) =$$

$$\sum_{c=1}^{G} \sum_{d=1}^{G} \sum_{a=1}^{K} \sum_{b=1}^{K} \lambda_c\, \lambda_d\, \beta_{G+a}\, \beta_{G+b}\ * \tag{2.30}$$

$$\sum_{k=0}^{t-i} \sum_{m=0}^{s-i} \rho_c^{t-i-k+1} \rho_d^{s-i-m+1} [\delta_{km} \sigma^2 E(x_{t,j} x_{s,j}) + \mathrm{cov}(x_{t,j} x_{k,a},\, x_{s,j} x_{m,b})]$$

By assumption 1, $|E(x_{t,j} x_{s,j})| \leq C + \mu_j^2$, so that

$$\left| \sum_{k=0}^{t-i} \sum_{m=0}^{s-i} \rho_c^{t-i-k+1} \rho_d^{s-i-m+1} \delta_{km} \sigma^2 E(x_{t,j} x_{s,j}) \right|$$

2.31

$$\leq \sigma^2 (C+\mu_j^2) \rho^2 (1-\rho^2)^{-1} \rho^{|t-s|}$$

Assumption 3 implies that

$$\left| \sum_{k=0}^{t-i} \sum_{m=0}^{s-i} \rho_c^{t-i-k+1} \rho_d^{s-i-m+1} \operatorname{cov}(x_{t,j}\, x_{k,a},\, x_{s,j}\, x_{m,b}) \right|$$

$$\leq C\pi^{|t-s|} \left\{ |t-s| \frac{\pi^2}{1-\pi^2} + \frac{\pi^2(3-\pi^2)}{(1-\pi^2)^2} \right\}$$

Lemmas 1.1, 1.2 may be invoked to show the convergence of this term. Therefore the convergence of the third term of 2.29 to zero uniformly in $\underline{\beta},\sigma^2$ is guaranteed.

The expected value of 2.9 is zero, since y_{t-i} is a linear combination of $\phi_1,\phi_2,\ldots,\phi_{t-i}$, all of which are independent of ε_t. The variance of 2.9 is

2.32
$$(n^2\sigma^6)^{-1} \sum_{t=1}^{i} y_{t-i}^2 + (n^2\sigma^4)^{-1} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \phi_k^2(t-i)$$

$$+ (n^2\sigma^6)^{-1} \sum_{t=i+1}^{n} \Phi_0^2(t-i)$$

$$+ (n^2\sigma^6)^{-1} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \sum_{\ell=1}^{t-i} \Phi_k(t-i)\, \Phi_\ell(t-i)\, E(m_k\, m_\ell)$$

Clearly the first and third terms of 2.32 go to zero uniformly in $\underline{\beta},\sigma^2$. Since $\sum_{k=1}^{t-i} \phi_k^2(t-i)$ is bounded above by a continuous function of $\underline{\beta},\sigma^2$, which is not a function of t,

$$n^{-2} \sum_{t=i+1}^{n} \sum_{k=1}^{t-i} \phi_k^2(t-i)$$

converges to zero uniformly for $\underline{\beta},\sigma^2$ in a compact set. Since

$$|E(m_k m_\ell)| \leq C(\sum_{a=1}^{K} \beta_{G+a})^2 + (\sum_{a=1}^{K} \beta_{G+a}\mu_a)^2$$

which is not a function of k or ℓ, we have that the last term of 2.32 is bounded in absolute value by

$$(n\sigma^6)^{-1}(\sum_{c=1}^{G} |\lambda_c|)^2 \rho^2(1-\rho)^{-2} [C(\sum_{a=1}^{K} \beta_{G+a})^2 + (\sum_{a=1}^{K} \beta_{G+a} \mu_a)^2]$$

which converges to zero as $n \to \infty$, uniformly for $\underline{\beta},\sigma^2$ in a compact set. Therefore the variance of 2.9 converges as required.

The expected value of 2.10 is zero, and the variance is, by assumption 1, bounded in absolute value by

$$(n\sigma^6)^{-1} (C + \mu_j)^2$$

which converges to zero as required.

Since the conditions for theorems 1.1 and 1.2 hold for 2.5 through 2.10, we apply theorem 1.3 to show the MLE of $\underline{\beta},\sigma^2$ have the desired properties. Theorem 2.2 is proved.

2.4 Optimality Results for the Multivariate Linear Model with Lagged Dependent Variables and Nonrandom Exogenous Variables

The results pertaining to optimal properties for the linear model

with lagged dependent variables may be extended to the situation where the variables are multivariate. The appropriate model is

$$2.33 \qquad \underline{y}_t = \sum_{g=1}^{G} B_g \underline{y}_{t-g} + \sum_{h=1}^{K^*} B_{G+h}\, \underline{x}_{t,h} + \underline{\varepsilon}_t \qquad t=1,2,\ldots$$

where $B_1,\ldots,B_G$ are the L by L matrices of unknown parameters and B_{G+h}, for $h = 1,\ldots,K^*$, are the L by M_h matrices of unknown parameters. The endogenous variable $\underline{y}_t$ is $L\times 1$ dimensioned. The exogenous variable $\underline{x}_{t,h}$ is $M_h \times 1$ dimensioned, for $h=1,\ldots,K^*$, and the error terms are $L\times 1$ dimensioned. To simplify the notation, define $\Gamma = [B_{G+1},\ldots,B_{G+K^*}]$ as the $L\times K$ dimensioned matrix of unknown coefficients of $\underline{z}_t^T = [\underline{x}_{t,1}^T, \underline{x}_{t,2}^T, \ldots, \underline{x}_{t,K^*}^T]$, which is $K\times 1$ dimensioned, where $K = \sum_{h=1}^{K^*} M_h$. The model 2.33 becomes

$$2.34 \qquad \underline{y}_t = \sum_{g=1}^{G} B_g\, \underline{y}_{t-g} + \Gamma \underline{z}_t + \underline{\varepsilon}_t \qquad t=1,2,\ldots$$

As in the univariate case, the stability condition will be assumed; i.e., all LG roots of the determinental equation

$$2.35 \qquad |\rho^G - \rho^{G-1} B_1 \,.\;.\;.\;. - B_G| = 0$$

are less than one in absolute value. Several theorems are presented below for different distributional assumptions on $\underline{z}_t$ and $\underline{\varepsilon}_t$.

For notational simplicity, the elements in row i, column j of a matrix A will be denoted by $(A)_{ij}$. We will also denote by $(\underline{v})_i$ the i'th component of the vector $\underline{v}$. The notation $(M)_{.k}$, $(M)_{k.}$ will denote

respectively the k'th column of M and the k'th row of M. The symbol σ^{ij} will denote the element in row i, column j of Σ^{-1}.

Theorem 2.3

For the model 2.34, assume that the exogenous variables $\underline{z}_t$ are nonrandom and the error terms are independent, with a joint normal distribution with mean vector $\underline{0}$ and nonsingular convariance matrix Σ. Assume also that

1) $|(\underline{z}_t)_h| \leq Z$ for $h=1,\ldots,K$, for $t=1,2,\ldots$

2) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n-c} \underline{z}_t \underline{z}_{t+c}^T$ exists, and the convergence is uniform in c, $c=1,2,\ldots$

Then the MLE of $B_1,\ldots,B_G,\Gamma,\Sigma$ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Consider the values of $\underline{y}_{1-G},\ldots,\underline{y}_0$ as constants. The conditional density of $\underline{y}_t$, given $\underline{y}_{t-1},\ldots,\underline{y}_{t-G}$ is multivariate normal with mean

$$\sum_{g=1}^{G} B_g \underline{y}_{t-g} + \Gamma \underline{z}_t$$

and with covariance matrix Σ. The joint density of $\underline{y}_1,\ldots,\underline{y}_n$ is therefore the product of the conditional densities at time t given the values of the G previous terms. The following second partial derivatives of the log likelihood function are required.

2.36 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial(B_g)_{ij}\,\partial(B_h)_{km}} = \sigma^{ki}\, n^{-1}\sum_{t=1}^{n} (\underline{y}_{t-g})_j\,(\underline{y}_{t-h})_m$$

$i,k,j,m = 1,\ldots,L;\ g,h = 1,\ldots,G.$

2.37 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial(\Gamma)_{ij}\,\partial(\Gamma)_{km}} = \sigma^{ki}\, n^{-1}\sum_{t=1}^{n} (\underline{z}_t)_j\,(\underline{z}_t)_m$$

$i,k = 1,\ldots,L;\ j,m = 1,\ldots,K$

2.38 $$\begin{aligned}-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ij}\,\partial\sigma_{hk}} &= -(\sigma^{hi}\sigma^{jk} + \sigma^{hj}\sigma^{ik})\\
&+ \sigma^{jk}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}\\
&+ \sigma^{ik}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}\\
&+ \sigma^{jh}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.k}\\
&+ \sigma^{ih}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.k}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}\end{aligned}$$

$i,j,h,k = 1,\ldots,L;\ i \neq j,\ h \neq k.$

2.39 $$\begin{aligned}-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ij}\,\partial\sigma_{hh}} &= -\sigma^{hi}\sigma^{jh}\\
&+ \sigma^{jh}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}\\
&+ \sigma^{ih}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}\end{aligned}$$

$i,j,h = 1,\ldots,L;\ i \neq j.$

2.40 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ii}\,\partial\sigma_{hh}} = -(1/2)\sigma^{ih}\sigma^{ih} + \sigma^{ih}\, n^{-1}\sum_{t=1}^{n} \underline{\varepsilon}_t^T (\Sigma^{-1})_{.h}\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.i}$$

$i,h = 1,\ldots,L$

2.41 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial(\Gamma)_{ij}\,\partial(B_g)_{km}} = \sigma^{ki}\, n^{-1}\sum_{t=1}^{n} (\underline{z}_t)_j\, (\underline{y}_{t-g})_m$$

$i,k,m = 1,\ldots,L;\; j = 1,\ldots,K$

2.42 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ij}\,\partial(B_g)_{km}} = \sigma^{jk}\, n^{-1}\sum_{t=1}^{n} (\underline{y}_{t-g})_m \underline{\varepsilon}_t^T (\Sigma^{-1})_{.i} + \sigma^{ik}\, n^{-1}\sum_{t=1}^{n} (\underline{y}_{t-g})_m\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.j}$$

$i \neq j;\; i,j,k,m = 1,\ldots,L;\; g = 1,\ldots,G$

2.43 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ii}\,\partial(B_g)_{km}} = \sigma^{ik}\, n^{-1}\sum_{t=1}^{n} (\underline{y}_{t-g})_m\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.i}$$

$i,k,m = 1,\ldots,L;\; g=1,\ldots,G.$

2.44 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ij}\,\partial(\Gamma)_{km}} = \sigma^{jk}\, n^{-1}\sum_{t=1}^{n} (\underline{z}_t)_m\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.i} + \sigma^{ik}\, n^{-1}\sum_{t=1}^{n} (\underline{z}_t)_m\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.j}$$

$i \neq j;\; i,j,k=1,\ldots,L;\; m=1,\ldots,K$

2.45 $$-n^{-1}\frac{\partial^2 \log L_n}{\partial\sigma_{ii}\,\partial(\Gamma)_{km}} = \sigma^{ik}\, n^{-1}\sum_{t=1}^{n} (\underline{z}_t)_m\, \underline{\varepsilon}_t^T (\Sigma^{-1})_{.i}$$

$i,k = 1,\ldots,L;\; m = 1,\ldots,K$

As in theorem 2.1, $\underline{y}_t$ can be expressed as

2.46 $$\underline{y}_t = \underline{\Phi}_0(t) + \Phi_1(t)\underline{\phi}_1 + \ldots + \Phi_t(t)\underline{\phi}_t \qquad t=1,\ldots$$

where $\underline{\phi}_i = \underline{\varepsilon}_i + \Gamma\underline{z}_i$, $\underline{\Phi}_0(t)$ is an $L\times 1$ vector of constants, and $\Phi_j(t)$, for $j=1,\ldots,t$, is an $L\times L$ matrix of constants. Generalizing from the proof of theorem 2.1, the system of equations which determine Φ is

2.47 $$\Phi_j(t) = B_1\Phi_j(t-1) + \ldots\ldots + B_G\Phi_j(t-G) \quad \text{for } j=0,\ldots,t$$

The initial conditions are

2.48 $$\Phi_j(j)=I,\ \Phi_j(j-1)=0,\ \ldots\ldots\ldots,\ \Phi_j(j-G+1)=0 \quad j=1,\ldots,n$$

$$\Phi_0(0)=\underline{y}_0,\ \Phi_0(-1)=\underline{y}_{-1},\ \ldots\ldots\ldots,\ \Phi_0(-G+1)=\underline{y}_{-G+1}$$

The matrices I and 0 are LxL. The solution to the equations 2.47 with initial conditions 2.48 is

2.49 $$\Phi_0(t) = \underline{Q}_1(t)\rho_1^t + \ldots\ldots\ldots + \underline{Q}_\ell(t)\rho_\ell^t$$

$$\Phi_j(t) = P_{1j}(t)\rho_1^t + \ldots\ldots + P_{\ell j}(t)\rho_\ell^t \quad j=1,2,\ldots,t$$

where $\rho_1,\rho_2,\ldots,\rho_\ell$ are the distinct solutions of 2.35. $\underline{Q}_1(t),\ldots,\underline{Q}_\ell(t)$ are Lx1 vectors whose elements are polynomial functions of t of degree equal to the multiplicity of the corresponding root, less one. $P_{ij}(t)$, for $i=1,\ldots,\ell$ for $j=1,\ldots,t$ are defined similarly. Assuming, without loss of generality, that all LG roots of 2.35 are

distinct, the vectors $\underset{\sim}{Q}$ and the matrices P have degree zero as polynomial functions of t. The solutions can therefore be written as

$$\Phi_j(t) = \sum_{k=1}^{LG} P_k \rho_k^{t-j+1} \qquad j=1,\ldots,t$$

$$\Phi_0(t) = \sum_{k=1}^{LG} \underline{Q}_k \rho_k^{t+1}$$

The expectations and variances of 2.36 through 2.45 can now be computed and shown to satisfy the assumptions of theorems 1.1 and 1.2.

For 2.36, if $h \geq g$

$$\begin{aligned} E(n^{-1} \sum_{t=1}^{n} (\underline{y}_{t-g})_j (\underline{y}_{t-h})_m) = \; & n^{-1} \sum_{t=1}^{g} (\underline{y}_{t-g})_j (\underline{y}_{t-h})_m + n^{-1} \sum_{t=g+1}^{h} (\underline{y}_{t-g})_j E(\underline{y}_{t-h})_m \\ & + n^{-1} \sum_{t=h+1}^{n} (\Phi_0(t-g))_j (\Phi_0(t-h))_m + n^{-1} \sum_{t=h+1}^{n} (\Phi_0(t-g))_j [\sum_{k=1}^{t-h} \Phi_k(t-h) E(\underline{\phi}_k)]_m \\ & + n^{-1} \sum_{t=h+1}^{n} (\sum_{k=1}^{t-g} \Phi_k(t-g) E(\underline{\phi}_k))_j (\Phi_0(t-h))_m \\ & + n^{-1} \sum_{t=h+1}^{n} \sum_{k=1}^{t-g} \sum_{m=1}^{t-h} (\Phi_k(t-g) E(\underline{\phi}_k \underline{\phi}_m^T) \Phi_m(t-h))_{jm} \end{aligned} \tag{2.50}$$

The first two terms of 2.50 clearly go to zero as $n \to \infty$, uniformly for $B_1,\ldots,B_G,\Gamma$ in a compact set. The third term is

$$\sum_{a=1}^{LG} \sum_{b=1}^{LG} (\underline{Q}_a)_j (\underline{Q}_b)_m (n^{-1} \sum_{t=h+1}^{n} \rho_a^{t-g+1} \rho_b^{t-h+1})$$

which converges to zero uniformly in $B_1,\ldots,B_G$, since ρ_a is a

continuous function of these parameters, and $0 \leq |\rho_a| < 1$, for $a=1,\ldots,G$. The fourth term of 2.50 is

$$\sum_{a=1}^{LG} \sum_{b=1}^{LG} (\underline{Q}_a)_j (n^{-1} \sum_{t=h+1}^{n} \sum_{k=1}^{t-h} \rho_a^{t-g+1} \rho_b^{t-h-k+1} \underline{z}_k^T) R_{bm}^T$$

where R_{bm} denotes the m'th row of $P_b\Gamma$. Since the elements of $\underline{z}_k$ are bounded above in absolute value, and $0 \leq |\rho_a| < 1$, with ρ_a a continuous function of $B_1,\ldots,B_G$, the above term goes to zero as $n \to \infty$ uniformly for $B_1,\ldots,B_G,\Gamma,\Sigma$ in a compact set. The same result holds for the fifth term. The sixth term is

$$\sum_{a=1}^{LG} \sum_{b=1}^{LG} R_{aj} (n^{-1} \sum_{t=h+1}^{n} (\sum_{k=1}^{t-g} \rho_a^{t-g-k+1} \sum_{m=1}^{t-h} \rho_b^{t-h-m+1} \underline{z}_k \underline{z}_m^T)) R_{bm}^T \tag{2.51}$$

$$+ \sum_{a=1}^{LG} \sum_{b=1}^{LG} (P_a \Sigma P_b^T)_{jm} (n^{-1} \sum_{t=h+1}^{n} \sum_{k=1}^{t-h} \rho_a^{t-g-k+1} \rho_b^{t-h-k+1})$$

Assumption 2 implies the first term of 2.51 converges, by the same argument used for the first term of 2.17. The second term of 2.51 converges to

$$\sum_{a=1}^{LG} \sum_{b=1}^{LG} (P_a \Sigma P_b^T)_{jm} \rho_a^{h-g} (\rho_a \rho_b)/(1-\rho_a \rho_b)$$

as $n \to \infty$, uniformly in the unknown parameters. Therefore we have shown that the expected value of 2.36 converges as $n \to \infty$, uniformly in the unknown parameters.

To verify that the variance of 2.36 goes to zero, we write it as

$$2.52 \quad n^2(\sigma^{ki})^2 \sum_{t=g+1}^{h} \sum_{s=g+1}^{h} \mathrm{cov}((\underline{y}_{t-g})_j, (\underline{y}_{s-g})_j) (\underline{y}_{t-h})_m (\underline{y}_{s-h})_m$$

$$+ 2(\sigma^{ki})^2 n^{-2} \sum_{t=g+1}^{h} \sum_{s=h+1}^{n} (\underline{y}_{t-h})_m \mathrm{cov}((\underline{y}_{t-g})_j, (\underline{y}_{s-g})_j (\underline{y}_{s-h})_m)$$

$$+ (\sigma^{ki})^2 n^{-2} \sum_{s=h+1}^{n} \sum_{t=h+1}^{n} \mathrm{cov}((\underline{y}_{t-g})_j (\underline{y}_{t-h})_m, (\underline{y}_{s-g})_j (\underline{y}_{s-h})_m)$$

The first two terms of 2.52 go to zero as $n \to \infty$, uniformly for the unknown parameters in a compact set, since each term is bounded above in absolute value by n^{-1} times a continuous function of the parameters. The double summation in the last term of 2.52 is analyzed in the same way as the corresponding univariate term, the last term of 2.19. Therefore we conclude that 2.52 goes to zero uniformly for the parameters in a compact set. This ends the proof that term 2.36 satisfies the assumptions of theorems 1.1 and 1.2.

Assumption 2 implies that term 2.37 converges as $n \to \infty$.

Terms 2.38, 2.39, 2.40 will be shown in section 3.2 to satisfy the conditions of theorems 1.1 and 1.2.

The expected value of 2.41 is the sum of a $o(1)$ term and

$$2.53 \qquad \sigma^{ik} \sum_{a=1}^{LG} (\underline{Q}_a)_m (n^{-1} \sum_{t=1}^{n} (\underline{z}_t)_j \rho_a^{t-h+1})$$

$$+ \sigma^{ik} \sum_{b=1}^{LG} R_{bm} \sum_{q=1}^{n-h} \rho_b{}^q (n^{-1} \sum_{t=h+q}^{n} (\underline{z}_t)_j \underline{z}_{t-h-q+1})$$

Term 2.53 is the multivariate analog of 2.23. The first term of 2.53 converges to zero as $n \to \infty$, since $(\underline{z}_t)_j$ is bounded by assumption 1,

and $0 \leq |\rho_a| < 1$. Assumption 2 implies the second term of 2.53 converges, applying a similar argument to that of the first term of 2.23. Again, the convergence is uniform for the parameters in a compact set.

The variance of 2.41 is

$$n^{-2}(\sigma^{ki})^2 \sum_{t=h+1}^{n} \sum_{s=h+1}^{n} (\underline{z}_t)_j (\underline{z}_s)_j \operatorname{cov}((\underline{y}_{t-h})_m, (\underline{y}_{s-h})_m).$$

For $s \geq t$, we have that $\operatorname{cov}((\underline{y}_{t-h})_m, (\underline{y}_{s-h})_m) =$

$$\sum_{a=1}^{LG} \sum_{b=1}^{LG} (P_a \Sigma P_b^T)_{mm} \rho_b^{s-t} (\rho_a \rho_b - (\rho_a \rho_b)^{t-h+1})(1-\rho_a \rho_b)^{-1}.$$

Since the components of $\underline{z}_t$, for all t, are bounded above in absolute value, and $\rho_j, j=1,\ldots,LG$ are less than one in absolute value, the variance of 2.41 goes to 0 as $n \to \infty$. The convergence is uniform in the parameters since ρ_a, P_a are continuous functions of the parameters.

To determine the expected value of 2.42, we note that $\underline{y}_{t-g}$ is a linear combination of error terms with indices less than or equal to $t-g$, and therefore independent of $\underline{\varepsilon}_t$. Therefore the expected value of this term is zero.

The variance of the first term of 2.42 may be bounded above by a continuous function of the unknown parameters which goes to zero as $n \to \infty$, uniformly in the parameters. The argument for 2.43 is similar.

The expected values of 2.44 and 2.45 are zero. The variances of both terms clearly go to zero, since the error terms are independent, and the components of the exogenous variables are bounded above in

absolute value.

Since the terms 2.36 through 2.45 satisfy the assumptions of theorems 1.1 and 1.2, we may apply theorem 1.3 to conclude that the hypotheses of Weiss's theorem are satisfied and, therefore the MLE of $B_1,\ldots,B_G,\Gamma,\Sigma$ have the desired properties.

2.5 Optimality Results for the Multivariate Linear Model with Lagged Dependent Variables and Random Exogenous Variables

Theorem 2.2 may also be extended to the multivariate model. The distributional assumptions for the components of the multivariate exogenous variables are the analogs of the distributional assumptions of the univariate theorem.

Theorem 2.4

For the model 2.34, when the stability condition is satisfied, assume that the exogenous variables $\{\underline{z}_t, t=1,2,\ldots\}$ are identically distributed, independently of the error sequence $\{\underline{\varepsilon}_t, t=1,2,\ldots\}$, which are independently distributed, multivariate normal with mean vector $\underline{0}$ and covariance matrix Σ. Assume also that assumptions 1,2,3 of theorem 2.2 hold when x is replaced by z. Then the MLE of $B_1,\ldots,B_G,\Gamma$, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Since the process $\underline{z}_t$ is independent of the error terms, and the parameters of its joint distribution are not functions of the parameters to be estimated, the second derivatives of the log likelihood function

with respect to the parameters are the same as 2.36 through 2.45. The arguments that these terms satisfy the assumptions of theorems 1.1 and 1.2 follow essentially the arguments of theorem 2.2 for the univariate model. The constants λ_m, m=1,...,G, are replaced in this theorem by the elements of the constant vectors $\underline{Q}_a$=1,...,LG or the elements of the constant matrices P_a,a=1,...,LG. The same arguments are used to show convergence and to bound the variances.

2.6 Linear Models with Autoregressive Disturbances

In theorems 2.1 through 2.4, it was assumed that the error terms were serially uncorrelated. The optimality properties also hold, under additional assumptions, when the errors satisfy a finite order autoregressive process. It is necessary to invoke theorem 1.4 in order to obtain the results. This approach is due to Nocturne [1970].

We will consider the multivariate autoregressive process 2.33, where the errors satisfy

$$\underline{\varepsilon}_t = R_1\underline{\varepsilon}_{t-1} + \cdots + R_H\underline{\varepsilon}_{t-H} + \underline{u}_t$$

where $\{\underline{u}_t, t=1,\ldots\}$ are distributed independently and identically, multivariate normal with mean $\underline{0}$ and covariance matrix Σ. Subtracting $R_1\underline{y}_{t-1} + R_2\underline{y}_{t-2} + \cdots + R_H\underline{y}_{t-H}$ from 2.33 yields

2.54
$$\underline{y}_t = \Delta_1\underline{y}_{t-1} + \cdots + \Delta_{G+H}\underline{y}_{t-G-H} + \theta_0\underline{z}_t + \cdots + \theta_H\underline{z}_{t-H} + \underline{u}_t \quad t=1,2,\ldots$$

where, assuming $G \geq H$,

$$\theta_0 = \Gamma$$

$$\theta_1 = -R_1\Gamma$$

$$\begin{array}{ll} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{array}$$

$$\theta_H = -R_H\Gamma$$

$$\Delta_1 = R_1 + B_1$$

$$\Delta_2 = R_2 + B_2 - R_1B_1$$

$$\Delta_3 = R_3 + B_3 - R_1B_1 - R_2B_1$$

$$\begin{array}{l} \cdot \\ \cdot \\ \cdot \end{array}$$

$$\Delta_{H+1} = B_{H+1} - R_1B_H - \cdot\ \cdot\ \cdot\ \cdot\ \cdot - R_HB_1$$

$$\begin{array}{l} \cdot \\ \cdot \end{array}$$

$$\Delta_{G+1} = -R_1B_G - R_2B_{G-1} - \cdot\ \cdot\ \cdot\ \cdot - R_HB_{G-H+1}$$

$$\begin{array}{l} \cdot \\ \cdot \\ \cdot \end{array}$$

$$\Delta_{G+H} = R_HB_G$$

We can apply theorems 2.3 or 2.4 to 2.54 to obtain optimal properties for the MLE of that model. The parameter map of theorem 1.4 is displayed above. It is clear that this function has the smoothness properties required. The map is 1-1 if there exist unique solutions $R_1,\ldots,R_H$ to the first H equations, when $\theta_j, j=1,\ldots,H$ are given. The equations for Δ_1 through Δ_G define $B_1,\ldots,B_G$. The parameters Δ_{G+1} through Δ_{G+H} are mapped to themselves. It should be noted that this method can be applied to any linear model with autoregressive disturbances.

CHAPTER III: NONLINEAR MODELS

3.1 Results from the Literature

Recent contributions have been made to the asymptotic theory of estimators for nonlinear models, and to computational methods for these estimators. Malinvaud [1966] deals with the nonlinear model

$$\underline{y}_t = \underline{g}(\underline{x}_t;\underline{\alpha}) + \underline{\varepsilon}_t \qquad t=1,2,\ldots \tag{3.1}$$

where $\underline{y}_t$ denotes the Lx1 dimensioned endogenous variable, $\underline{x}_t$ denotes the Kx1 dimensioned exogenous variable, $\underline{\alpha}$ denotes the pxl vector of unknown parameters, and $\underline{\varepsilon}_t$ denotes the Lx1 dimensioned error variable. Malinvaud chooses the estimator $\underline{\alpha}_n(S)$ which minimizes

$$\sum_{t=1}^{n} (\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}))^T S(\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}))$$

where S is a positive definite matrix. The following conditions are required to guarantee consistency and asymptotic normality for this estimator.

M1) The error variables $\{\underline{\varepsilon}_t\}$ are independently, identically distributed with mean vector $\underline{0}$ and nonsingular covariance matrix Σ.

M2) Let W be any closed set in Euclidean p-space such that the true value $\underline{\alpha}^0$ is in the complement of W. Let

$$Q_n(S,\underline{\alpha}) = \sum_{t=1}^{n} (\underline{g}(\underline{x}_t;\underline{\alpha})-\underline{g}(\underline{x}_t;\underline{\alpha}^0))^T S(\underline{g}(\underline{x}_t;\underline{\alpha})-\underline{g}(x_t;\underline{\alpha}^0))$$

As $n \to \infty$,

$$\sup_{\underline{\alpha} \text{ in } W} (Q_n(S,\underline{\alpha}))^{-1} \left| \sum_{t=1}^{n} (\underline{\varepsilon}_t)_h \; (\underline{g}(\underline{x}_t;\underline{\alpha}) - \underline{g}(\underline{x}_t;\underline{\alpha}^0))_i \right|$$

must converge stochastically to zero, for $i,h=1,\ldots,L$.

M3) An open neighborhood V of the true parameter value $\underline{\alpha}^0$ is contained in the set A of admissible values of $\underline{\alpha}$.

M4) In V, the functions $\underline{g}$ and their derivatives of the first three orders with respect to the unknown parameter are uniformly bounded in $\underline{x}_t$ and $\underline{\alpha}$. For any positive definite symmetric matrix S, the matrix

$$n^{-1} \sum_{t=1}^{n} Z_t^T \, S \, Z_t$$

where $(Z_t)_{ik} = \partial(\underline{g}(\underline{x}_t;\underline{\alpha})_i/\partial\alpha_k |_{\underline{\alpha}^0}$, is nonsingular and tends to a nonsingular matrix as $n \to \infty$.

Under assumptions M1 through M4, $\underline{\alpha}_n(S)$ is a consistent estimator for $\underline{\alpha}$ and $n^{1/2}(\underline{\alpha}_n(S) - \underline{\alpha}^0)$ is asymptotically normally distributed. Furthermore, if $\underline{\varepsilon}_t$ is normally distributed, the estimator $\underline{\alpha}_n(M_{\varepsilon\varepsilon}^{-1})$, where

$$M_{\varepsilon\varepsilon} = n^{-1} \sum_{t=1}^{n} (\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}_n(S)))(\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}_n(S)))^T$$

attains the Cramér-Rao lower bound on the variance. Under the same conditions, Barnett [1974] shows the MLE are consistent and asymptotically normally distributed, with the covariance matrix of the limiting distribution equal to the inverse of the information matrix.

Several authors have proven strong consistency for least squares estimators of parameters of nonlinear models. For the univariate model Jenrich [1969] uses 3.1, where $L=1$. The exogenous variables may be random or nonrandom. If they are assumed to be random, the limits in conditions J1, J2, and J3 below must exist almost surely, and the exogenous variables are independent of the error variables.

J1) $\lim_{n\to\infty} \sum_{t=1}^{n} (g(\underline{x}_t;\underline{\alpha}))^2$ must exist. Denote the limit as $g(\underline{\alpha})$. Furthermore, $|g(\underline{\alpha})-g(\underline{\alpha}^0)|^2$ has a unique minimum over $\underline{\alpha}$ in the compact set A at the true parameter value $\underline{\alpha}^0$.

Assumption J1 implies the least squares estimator $\hat{\underline{\alpha}}_n$ of $\underline{\alpha}^0$ and the estimator

$$\hat{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} (y_t - g(\underline{x}_t;\, \hat{\underline{\alpha}}_n))^2$$

of σ^2 converge almost surely to their true values.

J2) All first and second derivatives of $g(\underline{x}_t;\underline{\alpha})$ with respect to elements of $\underline{\alpha}$ must exist and be continuous on A. All limits of the form

$$n^{-1} \sum_{t=1}^{n} h_1(\underline{x}_t;\underline{\alpha})\, h_2(\underline{x}_t;\underline{\alpha})$$

exist, where h_1, h_2 are any of g, $g_i' = \partial g/\partial\alpha_i$, $g_{ij}'' = \partial^2 g/\partial\alpha_i\partial\alpha_j$, which must exist.

J3) The true parameter value $\underline{\alpha}^0$ must be an interior point of A, and the matrix whose i,j'th element is

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} g_i'(\underline{x}_t;\underline{\alpha})\ g_j'(\underline{x}_t;\underline{\alpha})\Big|_{\underline{\alpha}^0}$$

is nonsingular.

Under conditions J1, J2, J3, the least squares estimator of $\underline{\alpha}$ is asymptotically normally distributed with mean $\underline{0}$. The inverse of the asymptotic covariance matrix is given in condition J3. Furthermore, if normality of the error terms is assumed, then the least squares estimator is efficient in the Rao sense. A consistent estimator $\underline{T}_n$ of $\underline{\alpha}$ is said to be efficient in the Rao sense if

$$n^{1/2}|\underline{T}_n - \underline{\alpha}^0 - B\underline{z}_n|$$

converges stochastically to zero, where B is a matrix of constants which may depend on $\underline{\alpha}$, and $\underline{Z}_n$ is a vector whose i'th component is

$$n^{-1}\ \partial \log L_n(\underline{\alpha},\sigma^2)/\partial\alpha_i$$

Hannan [1971] extends Jennrich's result to the case where ε_t is generated by a stationary time series. Specifically, Hannan assumes ε_t is of the form

$$\varepsilon_t = \sum_{j=-\infty}^{\infty} \gamma_j \eta_{t-j} \qquad \text{where} \qquad \sum_{j=-\infty}^{\infty} (\gamma_j)^2 < \infty$$

The variables $\{\eta_t\}$ are assumed to be independently, identically distributed with mean 0 and variance 1. The following conditions are required.

H1) The spectrum of ε_t which is

$$f(\lambda) = (2\pi)^{-1} \left| \sum_{j=-\infty}^{\infty} \gamma_j \exp(ij\lambda) \right|^2$$

is continuous in λ. The set A of possible values of $\underline{\alpha}$ is a compact set.

H2) The exogenous variables $\{\underline{x}_t\}$ are independent of $\{\varepsilon_t\}$.

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} g(\underline{x}_t;\underline{\alpha}^1)\, g(\underline{x}_{t+k};\underline{\alpha}^2) = g(k,\underline{\alpha}^1,\underline{\alpha}^2)$$

exists almost surely for $k=0,\underline{+}1,\ldots$. The convergence is uniform in $\underline{\alpha}^1$ and $\underline{\alpha}^2$.

H3) $g(\underline{x};\underline{\alpha})$ is twice differentiable in $\underline{\alpha}$. Uniformly in $\underline{\alpha},\underline{\alpha}^1$, and $\underline{\alpha}^2$, the following limits exist almost surely, for $j,k=1,\ldots,p$ and for $c > 0$:

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} g_j'(\underline{x}_t;\underline{\alpha})\, g_k'(\underline{x}_{t+c};\underline{\alpha})$$

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} g_{jk}''(\underline{x}_t;\underline{\alpha}^1)\, g_{jk}''(\underline{x}_{t+c};\underline{\alpha}^2)$$

H4) $g(0,\underline{\alpha},\underline{\alpha}) + g(0,\underline{\alpha}^0,\underline{\alpha}^0) - 2g(0,\underline{\alpha},\underline{\alpha}^0) > 0$ for $\underline{\alpha} \neq \underline{\alpha}^0$.

Under H1, H2, and H4, if $\underline{\alpha}^0$ is an interior point of A, then the least squares estimator $\hat{\underline{\alpha}}_n$ of $\underline{\alpha}^0$ is consistent and asymptotically normally

distributed.

Robinson [1972] extends Hannan's results to the multivariate model

$$\underline{y}_t = B_0\, \underline{g}(\underline{x}_t;\underline{\alpha}^0) + \underline{\varepsilon}_t$$

where $\underline{\varepsilon}_t = \sum_{j=-\infty}^{\infty} \Gamma_j \underline{u}_{t-j}$, $\quad \sum_{j=-\infty}^{\infty} ||\Gamma_j||^2 < \infty$

where $\{\Gamma_j\}$ are LxL matrices with mean vector $\underline{0}$ and LxL identity covariance matrix. $||\Gamma_j||$ is defined as $[tr(\Gamma_j \Gamma_j^*)]^{1/2}$, where Γ_j^* is the adjoint of Γ_j. The parameters $\underline{\alpha}$ must satisfy s nonlinear equations

$$\begin{aligned} \xi_1(\underline{\alpha}) &= 0 \\ &\vdots \\ \xi_s(\underline{\alpha}) &= 0 \end{aligned}$$

which have continuous first derivatives. Under conditions which are the multivariate extensions of Hannan's conditions, the estimators $\hat{B}$ and $\hat{\underline{\alpha}}$ which minimize the residual sum of squares subject to the s nonlinear equations, are consistent and asymptotically normally distributed.

3.2 Optimality Results for Nonrandom Exogenous Variables and Independent Error Terms

We shall consider the multivariate model 3.1 under several different assumptions. In this section, the simplest assumptions are made. The exogenous variables are assumed to be nonrandom and the error terms are assumed to be independently distributed. For the

vector valued function $\underline{g}$ we use the notation

$$\underline{g}_i'(\underline{x}_t;\underline{\alpha}) = \partial\underline{g}(\underline{x}_t;\underline{\alpha})/\partial\alpha_i$$

$$\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha}) = \partial^2\underline{g}(\underline{x}_t;\underline{\alpha})/\partial\alpha_i\ \partial\alpha_j$$

Theorem 3.1

For model 3.1, assume the exogenous variables $\{\underline{x}_t, t=1,\ldots\}$ are nonrandom and the error variables $\{\underline{\varepsilon}_t, t=1,2,\ldots\}$ are independently, identically distributed, multivariate normal with mean vector $\underline{0}$ and unknown, positive definite covariance matrix Σ. Assume also that

1) $n^{-1}\sum_{t=1}^{n} \underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\Sigma^{-1}\underline{g}_j'(\underline{x}_t;\underline{\alpha})$ converges as $n\to\infty$, uniformly for $\underline{\alpha}$ in any compact set.

2) $|\underline{g}_i'(\underline{x}_t;\underline{\alpha})|$ is uniformly bounded in $\{\underline{x}_t\}$ and $\underline{\alpha}$, for $\underline{\alpha}$ in any compact set, for $i=1,\ldots,p$.

3) $|\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha})|$ is uniformly bounded in $\{\underline{x}_t\}$ and $\underline{\alpha}$, for $\underline{\alpha}$ in any compact set, for $i,j=1,\ldots,p$.

Then the MLE of $\underline{\alpha}$ and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

To apply theorems 1.1, 1.2, and 1.3, we calculate the second partial derivatives of the logarithm of the joint density of

$(\underline{y}_1,\ldots,\underline{y}_n)$, $L_n(\underline{\alpha},\Sigma)$.

3.2 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\sigma_{ij}\,\partial\sigma_{hk}} = -(\sigma^{hi}\sigma^{jk}+\sigma^{hj}\sigma^{ik})$$
$$+\sigma^{jk}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$
$$+\sigma^{ik}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}$$
$$+\sigma^{jh}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.k}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$
$$+\sigma^{ih}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.k}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}$$

for $i \neq j,\ h \neq k;\ i,j,h,k = 1,\ldots,L$.

3.3 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\sigma_{ij}\,\partial\sigma_{hh}} = -\sigma^{hi}\sigma^{hj}$$
$$+\sigma^{jh}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$
$$+\sigma^{ih}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}$$

for $i \neq j;\ i,j,h = 1,\ldots,L$.

3.4 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\sigma_{ii}\,\partial\sigma_{hh}} = -(1/2)\sigma^{hi}\sigma^{hi}$$
$$+\sigma^{hi}n^{-1}\sum_{t=1}^{n}\underline{\varepsilon}_t^T(\Sigma^{-1})_{.h}\,\underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$

for $i,h = 1,\ldots,L$

3.5 $$-n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\alpha_k\, \partial\sigma_{ij}} = n^{-1}\sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T (\Sigma^{-1})_{.i}\, \underline{\varepsilon}_t^T(\Sigma^{-1})_{.j}$$

$$+ n^{-1}\sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T (\Sigma^{-1})_{.j}\, \underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$

for $i \neq j$, $k = 1,\ldots,p$; $i,j = 1,\ldots,L$

3.6 $$-n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\alpha_k\, \partial\sigma_{ii}} = n^{-1}\sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T (\Sigma^{-1})_{.i}\, \underline{\varepsilon}_t^T(\Sigma^{-1})_{.i}$$

for $k = 1,\ldots,p$; for $i = 1,\ldots,L$.

3.7 $$-n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\Sigma)}{\partial\alpha_i\, \partial\alpha_h} = -n^{-1}\sum_{t=1}^{n} \underline{g}_{ih}''(\underline{x}_t;\underline{\alpha})^T \Sigma^{-1}\, \underline{\varepsilon}_t$$

$$+ n^{-1}\sum_{t=1}^{n} \underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\, \Sigma^{-1}\, \underline{g}_h'(\underline{x}_t;\underline{\alpha})$$

for $i,h = 1,\ldots,p$.

The expectations and variances of 3.2 through 3.7 will be shown to satisfy the assumptions of theorems 1.1, 1.2, and 1.3. The expected value of 3.2 is

3.8 $$\sigma^{jk}\,\sigma^{hi} + \sigma^{ik}\,\sigma^{hj}$$

and the variance of 3.2 is equal to n^{-1} times a continuous function of the elements of Σ, which goes to zero uniformly for Σ in a compact set. Terms 3.3 and 3.4 may be analyzed similarly.

The expectation of 3.5 is zero and the variance is

3.9 $$n^{-1}\,[\sigma^{jj}\, n^{-1} \sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T(\Sigma^{-1})_{.i}$$

$$+\ \sigma^{ii} n^{-1} \sum_{t=1}^{n} \underline{g}_k(\underline{x}_t;\underline{\alpha})^T(\Sigma^{-1})_{.j}$$

$$+\ 2\sigma^{ij}\, n^{-1} \sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T\,(\Sigma^{-1})_{.i}\, \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T(\Sigma^{-1})_{.j}]$$

By assumption 2, 3.9 goes to zero uniformly for $\underline{\alpha},\Sigma$ in a compact set. The argument for term 3.6 is identical.

The expectation of 3.7 is

3.10 $$n^{-1} \sum_{t=1}^{n} \underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\, \Sigma^{-1}\underline{g}_h'(\underline{x}_t;\underline{\alpha})$$

which, by assumption 1, converges as $n \to \infty$, uniformly for $\underline{\alpha},\Sigma$ in a compact set. Since $\underline{g}_i'$ is continuous as a function of $\underline{\alpha}$, the limit of 3.10 is continuous in $\underline{\alpha}$.

The variance of 3.7 is

$$n^{-1}[n^{-1} \sum_{t=1}^{n} \underline{g}_{ih}''(\underline{x}_t;\underline{\alpha})^T\ \Sigma^{-1}\ \underline{g}_{ih}''(\underline{x}_t;\underline{\alpha})]$$

which by assumption 3, converges to zero as $n \to \infty$, uniformly for $\underline{\alpha},\Sigma$ in a compact set.

The conditions of theorems 1.1, 1.2, 1.3 are satisfied; the MLE of $\underline{\alpha},\Sigma$ have the desired properties.

3.3 Optimality Results for Nonrandom Exogenous Variables and Error Terms which Form a Stationary Process

The independence restriction on the error terms may be relaxed. Suppose, specifically, that

$$\operatorname{cov}(\underline{\varepsilon}_t, \underline{\varepsilon}_s) = \rho^{|t-s|}\,\Sigma$$

where $|\rho| \leq 1$. Since the $\{\underline{\varepsilon}_t\}$ are normally distributed, they form a stationary process. The conditional distribution of $\underline{\varepsilon}_t$, given $\underline{\varepsilon}_{t-1},\ldots,\underline{\varepsilon}_1$ is

$$[2\pi(1-\rho^2)|\Sigma|]^{-1/2}\exp(-(2(1-\rho^2))^{-1}(\underline{\varepsilon}_t - \rho\underline{\varepsilon}_{t-1})^T\Sigma^{-1}(\underline{\varepsilon}_t - \rho\underline{\varepsilon}_{t-1}))$$

The joint density $L_n(\underline{\alpha},\rho,\Sigma)$ of $\underline{y}_1,\ldots,\underline{y}_n$ is the product over t from 2 to n of the conditional densities times the unconditional density of $\underline{\varepsilon}_1$.

Theorem 3.2

For the model 3.1, assume the variables $\{\underline{x}_t, t=1,2,\ldots\}$ are nonrandom and the error variables $\{\underline{\varepsilon}_t, t=1,2,\ldots\}$ form a multivariate stationary process whose density is multivariate normal with mean $\underline{0}$ and covariance function $\rho^{|s|}\Sigma$, where ρ,Σ are unknown except that $0 \leq \rho < 1$ and Σ is positive definite. Under assumptions 1,2,3 of theorem 3.1 and

4) $$n^{-1}\sum_{t=1}^{n} \underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\Sigma^{-1}\underline{g}_j'(\underline{x}_{t-1};\underline{\alpha}) \text{ converges, as } n \to \infty,$$

uniformly for $\underline{\alpha}, \Sigma$ in any compact set.

Then the MLE of $\underline{\alpha}, \rho, \Sigma$ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

The second derivatives of the log likelihood function are computed so that theorems 1.1, 1.2, and 1.3 may be applied. Let

$$\underline{u}_t = \underline{\varepsilon}_t - \rho\underline{\varepsilon}_{t-1} \qquad t=2,3,\ldots$$

$$\underline{u}_1 = \underline{\varepsilon}_1$$

Then the joint distribution of $\{\underline{u}_t, t=1,\ldots,n\}$ is multivariate normal with mean zero and

$$\text{cov}(\underline{u}_t, \underline{u}_s) = \delta_{ts}\,(1-\rho^2)\Sigma$$

The terms below will use this notation.

$$3.11 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\sigma_{ij}\,\partial\sigma_{hk}} = -\sigma^{hi}\sigma^{jk} - \sigma^{hj}\sigma^{ik}$$

$$+ \sigma^{jk} n^{-1} \underline{u}_1^T(\Sigma^{-1})_{.h}\, \underline{u}_1^T(\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1} \sigma^{jk} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T(\Sigma^{-1})_{.h} \underline{u}_t^T(\Sigma^{-1})_{.i}$$

$$+ n^{-1} \sigma^{ik} \underline{u}_1^T(\Sigma^{-1})_{.h}\, \underline{u}_1^T(\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1}\sigma^{ik} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T(\Sigma^{-1})_{.h} \underline{u}_t^T(\Sigma^{-1})_{.j}$$

$$+ \sigma^{jh} n^{-1} \underline{u}_1^T (\Sigma^{-1})_{.k} \underline{u}_1^T (\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1} \sigma^{jh} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T (\Sigma^{-1})_{.k} \underline{u}_t^T (\Sigma^{-1})_{.i}$$

$$+ \sigma^{ih} n^{-1} \underline{u}_1^T (\Sigma^{-1})_{.k} \underline{u}_1^T (\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1} \sigma^{ih} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T (\Sigma^{-1})_{.k} \underline{u}_t^T (\Sigma^{-1})_{.j}$$

for $i \neq j$; $h \neq k$, $i,j,h,k = 1,\ldots,L$.

3.12 $$-n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\sigma_{ij}\, \partial\sigma_{hh}} = - \sigma^{hi} \sigma^{hj}$$

$$+ \sigma^{jh} n^{-1} \underline{u}_1^T (\Sigma^{-1})_{.h} \underline{u}_1^T (\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1} \sigma^{jh} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T (\Sigma^{-1})_{.h} \underline{u}_t^T (\Sigma^{-1})_{.i}$$

$$+ \sigma^{ih} n^{-1} \underline{u}_1^T (\Sigma^{-1})_{.h} \underline{u}_1^T (\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1} \sigma^{ih} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T (\Sigma^{-1})_{.h} \underline{u}_t^T (\Sigma^{-1})_{.j}$$

for $i \neq j$; $i,j,h = 1,\ldots,L$

3.13 $$-n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\sigma_{ii}\, \partial\sigma_{hh}} = - (1/2)\sigma^{hi} \sigma^{hi} + \sigma^{hi} n^{-1} \underline{u}_1^T (\Sigma^{-1})_{.h} \underline{u}_1^T (\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1} \sigma^{hi} n^{-1} \sum_{t=2}^{n} \underline{u}_t^T (\Sigma^{-1})_{.h} \underline{u}_t^T (\Sigma^{-1})_{.i}$$

$i,h = 1,\ldots,L$

$$3.14 \quad -n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\alpha_k\,\partial\sigma_{ij}} = n^{-1}\underline{g}_k'(\underline{x}_1;\underline{\alpha})^T(\Sigma^{-1})_{.i}\underline{u}_1^T(\Sigma^{-1})_{.j}$$

$$+ n^{-1}\underline{g}_k'(\underline{x}_1;\underline{\alpha})^T(\Sigma^{-1})_{.j}\,\underline{u}_1^T(\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}[\underline{g}_k'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_k'(\underline{x}_{t-1};\underline{\alpha})]^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}[\underline{g}_k'(\underline{x}_t;\underline{\alpha}) - \rho\underline{g}_k'(\underline{x}_{t-1};\underline{\alpha})]^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.j}$$

$i \neq j;\ i,j = 1,\ldots,L;\ k = 1,\ldots,p.$

$$3.15 \quad -n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\alpha_k\,\partial\sigma_{ii}} = n^{-1}\underline{g}_k'(\underline{x}_1;\underline{\alpha})^T(\Sigma^{-1})_{.i}\underline{u}_1^T(\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}[\underline{g}_k'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_k'(\underline{x}_{t-1};\underline{\alpha})]^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.i}$$

for $i = 1,\ldots,L;\ k = 1,\ldots,p.$

$$3.16 \quad -n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\alpha_i\,\partial\alpha_j} = -\, n^{-1}\underline{g}_{ij}''(\underline{x}_1;\underline{\alpha})^T\Sigma^{-1}\underline{u}_1$$

$$+ n^{-1}\underline{g}_i'(\underline{x}_1;\underline{\alpha})^T\Sigma^{-1}\underline{g}_j'(\underline{x}_1;\underline{\alpha})$$

$$- (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}[\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_{ij}''(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}\underline{u}_t$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_j'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_j'(\underline{x}_{t-1};\underline{\alpha})]$$

for $i,j = 1,\ldots,p.$

$$3.17 \quad -n \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\rho\ \partial\sigma_{ij}} = -2\rho(1-\rho^2)^{-2}n^{-1}\sum_{t=2}^{n}\underline{u}_t^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.j}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T(\Sigma^{-1})_{.j}\underline{u}_t^T(\Sigma^{-1})_{.i}$$

for $i \neq j;\ i,j = 1,\ldots,L$

$$3.18 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\rho\ \partial\sigma_{ii}} = -\rho(1-\rho^2)^{-2}n^{-1}\sum_{t=2}^{n}\underline{u}_t^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.i}$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T(\Sigma^{-1})_{.i}\underline{u}_t^T(\Sigma^{-1})_{.i}$$

for $i = 1,\ldots L$

$$3.19 \quad -n^{-1} \frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial\rho\ \partial\alpha_i} = -2\rho(1-\rho^2)^{-2}n^{-1}\sum_{t=2}^{n}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}\underline{u}_t$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\Sigma^{-1}\underline{u}_t$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T\ \Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]$$

for $i = 1,\ldots,p$.

$$3.20 \quad -n^{-1}\frac{\partial^2 \log L_n(\underline{\alpha},\rho,\Sigma)}{\partial^2\rho} = -(n-1)n^{-1}(1+\rho^2)(1-\rho^2)^{-2}$$

$$+ (1+3\rho^2)(1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{u}_t^T\Sigma^{-1}\underline{u}_t$$

$$- 4\rho(1-\rho^2)^{-2}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T\Sigma^{-1}\underline{u}_t$$

$$+ (1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n}\underline{\varepsilon}_{t-1}^T\Sigma^{-1}\underline{\varepsilon}_{t-1}$$

To apply theorems 1.1, 1.2, and 1.3 to terms 3.11 through 3.20, the expected values of these terms will be shown to converge as $n \to \infty$ uniformly in the parameters, to a limit function, and the variances of these terms will be shown to converge, as $n \to \infty$, uniformly in the unknown parameters, to zero.

The expected value of 3.11 is equal to 3.8. To compute the variance term, we require the following.

$$\mathrm{cov}(\underline{u}_t^T(\Sigma^{-1})_{.h}\underline{u}_t^T(\Sigma^{-1})_{.i},\underline{u}_s^T(\Sigma^{-1})_{.k}\underline{u}_s^T(\Sigma^{-1})_{.j}) = \delta_{ts}(1-\rho^2)^2(\sigma_{hj}\sigma_{ik} + \sigma_{hk}\sigma_{ij})$$

Since all terms in the variance are zero unless $t=s$, the variance is a continuous function of the parameters which is $o(n^{-1})$. The assumptions of theorems 1.1 and 1.2 hold. By the same arguments the assumptions may be shown to hold for 3.12 and 3.13.

The expectation of 3.14 is zero, since the $\{\underline{u}_t,t=1,\ldots\}$ have zero expectation. We will compute one of the terms in the variance of 3.14, the covariance of the second and third terms. This covariance is

$$(1-\rho^2)^{-1}\sigma_{ij}n^{-2}\sum_{t=1}^{n}[\underline{g}_k(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_k(\underline{x}_{t-1};\underline{\alpha})]^T(\Sigma^{-1})_{.i}[\underline{g}_k'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_k'(\underline{x}_{t-1};\underline{\alpha})]^T(\Sigma^{-1})_{.j}$$

By assumption 2, this term converges to zero as $n \to \infty$, uniformly for the parameters in any compact set. The other terms of the variance of 3.14 converge to zero by the same argument. Thus the assumptions of theorems 1.1 and 1.2 are satisfied for 3.14 and 3.15.

The expected value of 3.16 is

$$n^{-1} \underline{g}_i'(\underline{x}_1;\underline{\alpha})^T \Sigma^{-1} \underline{g}_j'(\underline{x}_1;\underline{\alpha})$$
$$+ (1-\rho^2)^{-1} n^{-1} \sum_{t=2}^{n} [\underline{g}_i'(\underline{x}_t;\underline{\alpha}) - \rho \underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T \Sigma^{-1} [\underline{g}_j'(\underline{x}_t;\underline{\alpha}) - \rho \underline{g}_j'(\underline{x}_{t-1};\underline{\alpha})]$$

By assumptions 1 and 4, this term converges as $n \to \infty$, uniformly for $\underline{\alpha}, \Sigma, \rho$ in a compact set. To compute the variance of 3.16, we note that the variance of the third term is

$$(1-\rho^2)^{-1} n^{-2} \sum_{t=2}^{n} [\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha}) - \rho \underline{g}_{ij}''(\underline{x}_{t-1};\underline{\alpha})]^T \Sigma^{-1} [\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha}) - \rho \underline{g}_{ij}''(\underline{x}_{t-1};\underline{\alpha})]$$

By assumption 3, this term goes to zero as $n \to \infty$, uniformly for the parameters in a compact set. Therefore, by assumption 3, the variance of 3.16 satisfies the conditions of theorem 1.1.

To compute the expected value of 3.17, we note that

$$E(\underline{\varepsilon}_{t+k})(\underline{\varepsilon}_t - \rho\, \underline{\varepsilon}_{t-1}) = \rho^k (1-\rho^2)\Sigma \qquad \text{for } k \geq 0$$

$$E(\underline{\varepsilon}_{t-m})(\underline{\varepsilon}_t - \rho\, \underline{\varepsilon}_{t-1}) = 0 \qquad \text{for } m \geq 1$$

The expected value of 3.17 is

$$-2\rho(1-\rho^2)^{-1}(n-1)n^{-1}\sigma_{ij}$$

which satisfies the conditions of theorem 1.1. To compute the variance of 3.17, we note the following. The variance of the first term of 3.17 is

$$4\rho^2(1-\rho^2)^{-2}(1-n^{-1})n^{-1}(\sigma_{ii}\sigma_{jj}+\sigma_{ij}^2)$$

To compute the variance of the second term of 3.17, note that

$$\mathrm{cov}((\underline{\varepsilon}_{t-1})_a(\underline{u}_t)_b,(\underline{\varepsilon}_{s-1})_c(\underline{u}_s)_d)=\delta_{ts}(1-\rho^2)\sigma_{ac}\sigma_{bd}$$

so that the variance of the second term is

$$(1-\rho^2)^{-1}n^{-1}\sigma_{ii}\sigma_{jj}$$

The variance of the third term is computed similarly to the variance of the second term. To compute the covariance of the first and second terms of 3.17, we note that

$$\mathrm{cov}((\underline{u}_t)_a(\underline{u}_t)_b,(\underline{\varepsilon}_{s-1})_c(\underline{u}_s)_d)=0$$

Therefore the covariance of the first and second terms of 3.17 is zero. By the same argument the covariance of the first and third terms of 3.17 is zero. Finally we conclude that the variance of 3.17 converges to zero uniformly for the parameters in a compact set. Term 3.18 can

be shown to satisfy the conditions of theorems 1.1 and 1.2 by the same argument.

The expected value of 3.19 is zero, since the $\{\underline{u}_t\}$ have expectation zero. The variance of the first term of 3.19 is

$$4\rho^2(1-\rho^2)^{-3}n^{-2}\sum_{t=2}^{n}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]$$

which, by assumption 2, converges to zero uniformly for the parameters in a compact set. Similarly, the variance of the second term of 3.19 goes to zero. The variance of the third term of 3.19 is

$$(1-\rho^2)^{-2}n^{-2}\sum_{t=2}^{n}\sum_{s=2}^{n}\rho^{|t-s|}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]$$

which, by assumption 2, also converges to zero as required. The covariance of the second and third terms of 3.19 is

$$(1-\rho^2)^{-1}n^{-2}\sum_{t=2}^{n-1}\sum_{s=t+1}^{n}\rho^{s-1-t}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]$$

which, by assumption 2, converges to zero uniformly for the parameters in a compact set. The covariance of the first and third terms of 3.19 behaves similarly. Thus we may conclude that 3.19 satisfies the conditions of theorems 1.1 and 1.2.

The expected value of 3.20 is

$$-(1-n^{-1})[(1+\rho^2)(1-\rho^2)^{-2} - (1+3\rho^2)L - (1-\rho^2)^{-1}L]$$

which clearly converges as $n \to \infty$, uniformly for ρ in a compact set in $[0,1)$. The variance of the second term of 3.20 is

$$2(1+3\rho^2)^2 (1-n^{-1})n^{-1} \sum_{b=1}^{L} \sum_{d=1}^{L} (\sigma_{bd})^2$$

which obviously goes to zero as $n \to \infty$, uniformly for values of ρ, Σ in a compact set. The variance of the third term is

$$16\rho^2(1-\rho^2)^{-1} (1-n^{-1})n^{-1} \sum_{b=1}^{L} \sum_{d=1}^{L} (\sigma_{bd})^2$$

which goes to zero as required. The variance of the fourth term of 3.20 is

$$2(1-\rho^2)^{-2}[n^{-2} \sum_{t=2}^{n} \sum_{s=2}^{n} \rho^{2|t-s|}] \sum_{b=1}^{L} \sum_{d=1}^{L} (\sigma_{bd})^2$$

which goes to zero as required. By the same argument which proved that the covariance of the first and second terms of 3.17 was zero, the covariance of the second and third terms of 3.20 is zero. The covariance of the third and fourth terms of 3.20 is

$$-4\rho(1-\rho^2)^{-2} \{2 \sum_{b=1}^{L} \sum_{d=1}^{L} (\sigma_{bd})^2\} n^{-2} \sum_{t=2}^{n-1} \sum_{s=t+1}^{n} \rho^{s-1-t}$$

which goes to zero as $n \to \infty$, uniformly for the parameters in a compact set. The covariance of the second and fourth terms of 3.20 is

$$(1+3\rho^2) \{2 \sum_{b=1}^{L} \sum_{d=1}^{L} (\sigma_{bd})^2\} n^{-2} \sum_{t=2}^{n-1} \sum_{s=t+1}^{n} \rho^{s-1-t}$$

which converges to zero as required.

We have shown that the conditions of theorems 1.1 and 1.2 are satisfied for terms 3.11 through 3.20. Applying theorem 1.3, we conclude that the MLE of the parameters have the properties required.

3.4 Optimality Results for Random Exogenous Variables and Independent Error Terms

Optimality properties may also be obtained for the model 3.1 when the exogenous variables are random. Assuming that the $\{\underline{x}_t\}$ are a stationary process is not sufficient to obtain the results, since the stochastic convergence of functions of $\underline{g}_i'(\underline{x}_t;\underline{\alpha})$ and $\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha})$, which are not necessarily stationary, is required.

<u>Theorem 3.3</u>

For model 3.1, assume the variables $\{\underline{x}_t\}$ are random and identically distributed, independently of the error terms $\{\underline{\varepsilon}_t\}$, which are independently distributed, multivariate normal with mean $\underline{0}$ and positive definite covariance Σ. Assume also that

1) $|E(\underline{g}_i'(\underline{x}_s;\underline{\alpha})\underline{g}_j'(\underline{x}_t;\underline{\alpha})^T)|$ exists, and all elements are bounded above uniformly for $\underline{\alpha}$ in any compact set, and for $t,s = 1,2,\ldots$
2) $|E(\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha})\underline{g}_{ij}''(\underline{x}_s;\underline{\alpha})^T)|$ exists, and all elements are bounded above uniformly for $\underline{\alpha}$ in any compact set, and for $t,s = 1,2,\ldots$
3) There exist constants $C > 0$ and $0 \leq \eta < 1$, such that

for $i,j = 1,\ldots,p$ and for $t,s = 1,2,\ldots$

$$|\mathrm{cov}(\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\ \Sigma^{-1}\underline{g}_j'(\underline{x}_t;\underline{\alpha}),\ \underline{g}_i'(\underline{x}_s;\underline{\alpha})^T\ \Sigma^{-1}\underline{g}_j'(\underline{x}_s;\underline{\alpha}))| < C\eta^{|t-s|}$$

for all $\underline{\alpha}$ in any compact set.

Then the MLE of $\underline{\alpha}$ and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

The second partial derivatives of the log likelihood function are terms 3.2 through 3.7, since the $\{\underline{x}_t\}$ are distributed independently of $\{\underline{\varepsilon}_t\}$, and their common density is not a function of the unknown parameters. The expectation and variance of terms 3.2, 3.3, and 3.4 are as calculated in the proof of theorem 3.1.

The expected value of term 3.5 is zero and the variance is

$$n^{-1}\{\sigma^{jj}(\Sigma^{-1})_{i.}\ E(\underline{g}_k'(\underline{x};\underline{\alpha})\underline{g}_k'(\underline{x};\underline{\alpha})^T)(\Sigma^{-1})_{.i}$$

$$+\ 2\sigma^{ij}(\Sigma^{-1})_{j.}\ E(\underline{g}_k'(\underline{x};\underline{\alpha})\underline{g}_k'(\underline{x};\underline{\alpha})^T)(\Sigma^{-1})_{.i}$$

$$+\ \sigma^{ii}\ (\Sigma^{-1})_{j.}\ E(\underline{g}_k'(\underline{x};\underline{\alpha})\underline{g}_k'(\underline{x};\underline{\alpha})^T)(\Sigma^{-1})_{.j}\}$$

By assumption 1, the term in brackets above is bounded uniformly for $\underline{\alpha}$, Σ in a compact set. Therefore the conditions of theorems 1.1 and 1.2 hold for 3.5. The mean and variance of 3.6 are computed similarly.

The expected value of 3.7 is

$$E(\underline{g}_i'(\underline{x};\underline{\alpha})^T\ \Sigma^{-1}\ \underline{g}_h'(\underline{x};\underline{\alpha}))$$

which exists by assumption 1 and is continuous as a function of $\underline{\alpha}$ since $\underline{g}_{ij}''$ exists. The variance is

$$n^{-1} E(\underline{g}_{ih}''(\underline{x};\underline{\alpha})^T \Sigma^{-1} \underline{g}_{ih}''(\underline{x};\underline{\alpha}))$$

$$+ n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} \operatorname{cov}(\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T \Sigma^{-1} \underline{g}_h'(\underline{x}_t;\underline{\alpha}),\ \underline{g}_i'(\underline{x}_s;\underline{\alpha})^T \Sigma^{-1} \underline{g}_h'(\underline{x}_s;\underline{\alpha}))$$

By assumption 2, the first term goes to 0 as $n \to \infty$ uniformly for $\underline{\alpha},\Sigma$ in a compact set. By assumption 3, the uniform convergence of the second term is assured. Therefore theorem 1.3 may be applied to guarantee the optimal properties.

3.5 Optimality Properties for Random Exogenous Variables Independent of the Error Variables, which Form a Stationary Process

We are able to prove optimal properties for the MLE when the sequences $\{\underline{x}_t\}$ and $\{\underline{\varepsilon}_t\}$ are independent, but the variables within each sequence may be dependent, with certain restrictions. The restrictions that the $\{\underline{x}_t\}$ be identically distributed is removed, since it does not greatly simplify the assumptions.

Theorem 3.4

For the model 3.1, assume the variables $\{\underline{x}_t\}$ are random and distributed independently of the error variables $\{\underline{\varepsilon}_t\}$ which form a multivariate stationary process. The common distribution of each $\{\underline{\varepsilon}_t\}$ is multivariate normal with mean vector $\underline{0}$. The covariance matrix of $\underline{\varepsilon}_t$ and $\underline{\varepsilon}_s$ is $\rho^{|t-s|}\Sigma$, for $t,s = 1,2,\ldots$, where ρ and

Σ are unknown except that $0 \leq \rho < 1$ and Σ is positive definite. Assume also that

1) $\lim_{n\to\infty} n^{-1} \sum_{t=2}^{n} E[\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T \Sigma^{-1} \underline{g}_j'(\underline{x}_s;\underline{\alpha})]$ exists for $s=t$ or $s=t-1$. For both limits the convergence is uniform for $\underline{\alpha}$ in any compact set.

2) The elements of $|E[\underline{g}_i'(\underline{x}_t;\underline{\alpha})\underline{g}_j'(\underline{x}_s;\underline{\alpha})^T]$ are bounded above uniformly for $\underline{\alpha}$ in any compact set and for $s,t = 1,2,\ldots$

3) The elements of $|E[\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha})\underline{g}_{ij}''(\underline{x}_s;\underline{\alpha})^T]|$ are bounded above uniformly for $\underline{\alpha}$ in compact set and for $s,t = 1,2,\ldots$

4) $|\mathrm{cov}[\underline{g}_i'(\underline{x}_u;\underline{\alpha})^T \Sigma^{-1} \underline{g}_j'(\underline{x}_u;\underline{\alpha}),\ \underline{g}_i'(\underline{x}_v;\underline{\alpha})^T \Sigma^{-1} \underline{g}_j'(\underline{x}_v;\underline{\alpha})]| \leq Cr^{|t-s|}$ for $u = t,\ t-1$, for $v = s,\ s-1$ and for values of $\underline{\alpha}$ in any compact set, where $C > 0$ and $0 \leq r < 1$.

Then the MLE of $\underline{\alpha}, \rho$, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Since the variables $\{\underline{x}_t\}$ are distributed independently of the error terms, and their joint density is not a function of the unknown parameters, the second partial derivatives of the log likelihood are as computed in 3.11 through 3.20 of theorem 3.2.

The arguments that 3.11, 3.12, and 3.13 satisfy the conditions of theorems 1.1 and 1.2 are identical to the arguments in the proof of theorem 3.2 since these terms do not involve the exogenous variables.

By assumption 2 and the independence of $\{\underline{\varepsilon}_t\}$ and $\{\underline{x}_t\}$, the expected value of 3.14 is zero. The covariance of the third and fourth terms of 3.14 is

$$\sigma_{ij}(1-\rho^2)n^{-2}\sum_{t=2}^{n} h_k(\underline{x}_t;\underline{\alpha})^T(\Sigma^{-1})_{.i}\ h_k(\underline{x}_t;\underline{\alpha})^T(\Sigma^{-1})_{.j}$$

$$\text{where}\quad h_k(\underline{x}_t;\underline{\alpha}) = \underline{g}_k'(\underline{x}_t;\underline{\alpha}) - \rho\underline{g}_k'(\underline{x}_{t-1};\underline{\alpha})$$

By assumption 2, the above term goes to zero as $n \to \infty$ uniformly in a compact set of the parameters. The same argument applies to all terms in the variance of 3.14. The expectation and variance of 3.15 behave similarly.

By assumptions 2 and 3, and the independence of $\{\underline{x}_t\}$ and $\{\underline{\varepsilon}_t\}$, the expected value of 3.16 exists and is equal to

$$n^{-1}\,E(\underline{g}_i'(\underline{x}_1;\underline{\alpha})^T\,\Sigma^{-1}\,\underline{g}_j'(\underline{x}_1;\underline{\alpha})) +$$

$$(1-\rho^2)^{-1}n^{-1}\sum_{t=2}^{n} E\{[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\,\Sigma^{-1}[\underline{g}_j'(\underline{x}_t;\underline{\alpha}) - \rho\underline{g}_j'(\underline{x}_{t-1};\underline{\alpha})]\}$$

By assumption 1, the expected value of 3.16 converges as $n \to \infty$ uniformly in a compact set of the parameters. To compute the variance of 3.16, we note that the variance of the third term is

$$(1-\rho^2)n^{-2}\sum_{t=2}^{n} E[\underline{g}_{ij}''(\underline{x}_t;\underline{\alpha}) - \rho\underline{g}_{ij}''(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_{ij}''(\underline{x}_t;\alpha) - \rho\underline{g}_{ij}''(\underline{x}_{t-1};\underline{\alpha})]$$

which, by assumption 3, converges to zero in the manner required. Assumption 4 implies that the fourth term of 3.16 behaves similarly. Since the covariance of the third and fourth terms of 3.16 is zero,

we may conclude that the variance of 3.16 converges to zero uniformly in a compact set of the parameters.

Since terms 3.17 and 3.18 are not functions of $\{\underline{x}_t\}$, their expectations and variances converge as in the proof of theorem 3.2.

The independence of the sequences $\{\underline{x}_t\}$ and $\{\underline{\varepsilon}_t\}$ and assumption 2 imply that the expected value of 3.19 is zero. The variance of the first term of 3.19 is

$$4\rho^2(1-\rho^2)^{-1}n^{-2}\sum_{t=2}^{n} E\{[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]\}$$

which, by assumption 2, converges to zero in a compact set of the parameters. Assumption 2 implies that the variance of the second term behaves similarly. The variance of the third term is

$$(1-\rho^2)^{-2}n^{-2}\sum_{t=2}^{n}\sum_{s=2}^{n}\rho^{|t-s|} *$$

$$E\{[\underline{g}_i'(\underline{x}_t;\underline{\alpha})-\rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\ \Sigma^{-1}[\underline{g}_i'(\underline{x}_s;\underline{\alpha}) - \rho\underline{g}_i'(\underline{x}_{s-1};\underline{\alpha})]\}$$

Lemma 1.1 and assumption 2 imply that the above term goes to zero as required. The covariance of the second and third terms of 3.19 is

$$(1-\rho^2)^{-1}n^{-2}\sum_{t=2}^{n-1}\sum_{s=t+1}^{n}\rho^{\,s-1-t} *$$

$$E\{[\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]^T\ \Sigma^{-1}[\underline{g}_i'(\underline{x}_t;\underline{\alpha}) - \rho\underline{g}_i'(\underline{x}_{t-1};\underline{\alpha})]\}$$

which, by assumption 2, converges to zero as required. By the same

argument, the covariance of the first and third terms of 3.19 converges to zero as $n \to \infty$. Therefore the variance of 3.19 satisfies the conditions of theorems 1.1 and 1.2.

By the arguments of theorem 3.2, the expectation and variance of 3.20 satisfy the conditions of theorems 1.1 and 1.2.

Theorem 1.3 implies that the MLE of $\underline{\alpha}, \rho$ and Σ have the properties required.

3.6 Optimal Properties when Independence Between Exogenous Variables and Error Variables is not Assumed

When independence between the exogenous variables and the error terms is not assumed, the joint conditional distribution of $\underline{\varepsilon}_1,\ldots,\underline{\varepsilon}_n$ given $\underline{x}_1,\ldots,\underline{x}_n$ must be known in order to obtain the MLE of the unknown parameters. For simplicity we are assuming that the distributional parameters of the marginal density of $\underline{x}_1,\ldots,\underline{x}_n$ are not functions of the parameters to be estimated. Also, we will assume that the joint distribution of $(\underline{\varepsilon}_1,\ldots,\underline{\varepsilon}_n,\underline{x}_1,\ldots,\underline{x}_n) = (\underline{\varepsilon}(n),\underline{x}(n))$ is multivariate normal, of dimension $n(L+K)$ with mean vector $\underline{0}$ and covariance matrix

3.21
$$\Sigma = \begin{bmatrix} \Sigma_\varepsilon & \Sigma_{\varepsilon x} \\ \Sigma_{x\varepsilon} & \Sigma_x \end{bmatrix}$$

where the covariance matrix is partitioned conformably with $\underline{\varepsilon}(n)$ and $\underline{x}(n)$. The joint conditional distribution of the error variables, given the exogenous variables, is multivariate normal of dimension nL with mean

AD-A042 592 CORNELL UNIV ITHACA N Y SCHOOL OF OPERATIONS RESEARC--ETC F/G 5/3
OPTIMAL ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS --ETC(U)
MAY 77 M K VICKERS DAAG29-73-C-0008

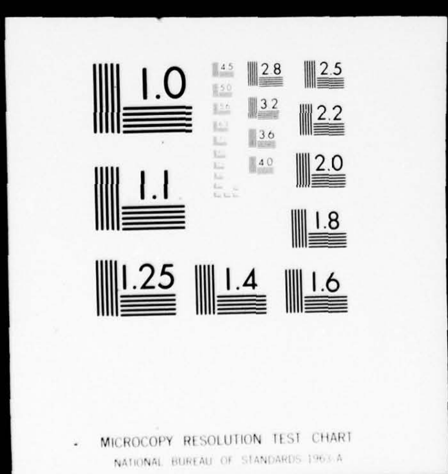
UNCLASSIFIED TR-334 NL

2 OF 2
ADA042592

END
DATE
FILMED
8-77
DDC

2 OF 2

ADA042592



$$\underline{c}(n) = \Sigma_{\varepsilon x} \Sigma_x^{-1} \underline{x}(n)$$

and covariance matrix

3.22 $$\Omega(n) = \Sigma_\varepsilon - \Sigma_{\varepsilon x}\Sigma_x^{-1} \Sigma_{x\varepsilon}.$$

The conditional distribution of $\underline{y}_1,\ldots,\underline{y}_n$ given $\underline{x}(n)$ is multivariate normal with mean $\underline{g}(n;\underline{\alpha}) + \underline{c}(n)$, where

$$\underline{g}(n;\underline{\alpha})^T = (\underline{g}(\underline{x}_1;\underline{\alpha})^T, \ . \ . \ . \ , \underline{g}(\underline{x}_n;\underline{\alpha})^T).$$

The covariance matrix is $\Omega(n)$. For future reference, the t'th component vector of $\underline{c}(n)$ will be denoted $\underline{c}_t$, $t=1,\ldots,n$.

Without the assumption that $\{\underline{\varepsilon}_t\}$ and $\{\underline{x}_t\}$ are jointly normally distributed, the conditional distribution is difficult to obtain. The assumption that the exogenous variables have $\underline{0}$ mean is made so that the MLE of $\underline{\alpha}$ is not a function of this unknown mean. The exogenous variables are not required to be identically distributed. We will estimate the conditional covariance matrix $\Omega(n)$ of the error terms, under two assumptions. It is not necessary to assume that the error variables are identically distributed or that they are independent. Instead, the joint distribution of $\underline{\varepsilon}(n)$ and $\underline{x}(n)$ will be assumed to have the following properties:

1) $\text{var}(\underline{\varepsilon}_t-\underline{c}_t) = \text{var}(\underline{\varepsilon}_s-\underline{c}_s)$ for $t,s=1,2,\ldots$
2) $\text{cov}(\underline{\varepsilon}_t,\underline{\varepsilon}_s) = \text{cov}(\underline{\varepsilon}_t,\underline{c}_s) = \text{cov}(\underline{c}_t,\underline{\varepsilon}_s) = \text{cov}(\underline{c}_t,\underline{c}_s)$ for $t\neq s$, $t,s=1,2,\ldots$

The vector $\underline{c}(n)$ is that vector in the linear subspace spanned by $\underline{x}(n)$ which maximizes the covariance with $\underline{\varepsilon}(n)$. Stated differently, the regression of $\underline{\varepsilon}(n)$ on $\underline{x}(n)$ produces $\Sigma_{\varepsilon x}\Sigma_{x}^{-1}$ as the vector of coefficients. The vector difference $\underline{\varepsilon}(n) - \underline{c}(n)$ is uncorrelated with $\underline{x}(n)$. The second assumption above implies that the covariance between different disturbance terms is completely attributable to variation in their projection on the subspace spanned by $\underline{x}(n)$. The first condition assures the homoscedasticity of that part of the error term which is not attributable to $\underline{x}(n)$. Under these assumptions, the covariance matrix $\Omega(n)$ is block diagonal; the covariance matrices which form the blocks are equal, say, to Ω. The log likelihood function is

3.23 $$- (n/2) \log|\Omega| - (1/2)[\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}) - \underline{c}_t]^T \Omega^{-1} [\underline{y}_t - \underline{g}(\underline{x}_t;\underline{\alpha}) - \underline{c}_t]$$

plus a term which is not a function of the unknown parameters.

<u>Theorem 3.5</u>

For the model 3.1, assume that the joint distribution of $\underline{\varepsilon}(n)$ and $\underline{x}(n)$ is multivariate normal with mean $\underline{0}$ and covariance matrix given by 3.21. Assume that conditions 1 and 2 above hold for the covariance matrix. Assume also that

3) The elements of $|E(h(\underline{x}_t;\underline{\alpha})h(\underline{x}_t;\underline{\alpha})^T)|$ are bounded above for $t=1,2,\ldots$ and for $\underline{\alpha}$ in any compact set, where the function h is one of $\underline{g}_k'$ or $\underline{g}_{ij}''$. Also, the elements of $|E(\underline{g}_k'(\underline{x}_t;\underline{\alpha})\underline{g}_{ij}''(\underline{x}_s;\underline{\alpha})^T)|$ are bounded above for $t,s=1,2,\ldots,$ for $\underline{\alpha}$ in any compact set.

4) $n^{-1}\sum_{t=1}^{n} E(\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\Omega^{-1}\underline{g}_j'(\underline{x}_s;\underline{\alpha}))$ converges as $n \to \infty$, uniformly for $\underline{\alpha},\Omega$ in any compact set, for $i,j=1,\ldots,p$, for $t,s=1,2,\ldots$

5) $|\text{cov}[\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T\Omega^{-1}\underline{g}_j'(\underline{x}_t;\underline{\alpha}),\ \underline{g}_i'(\underline{x}_s;\underline{\alpha})^T\Omega^{-1}\underline{g}_j'(\underline{x}_s;\underline{\alpha})]| \leq C\eta^{|t-s|}$ for $i,j=1,\ldots,p$, for $t,s=1,2,\ldots$, where $C > 0$ and $0 \leq \eta < 1$.

6) $n^{-1}\sum_{t=1}^{n} E[h(\underline{x}_t;\underline{\alpha})\underline{\varepsilon}_t^T]$, $n^{-1}\sum_{t=1}^{n} E[h(\underline{x}_t;\underline{\alpha})\underline{c}_t^T]$ converge, as $n \to \infty$, uniformly for $\underline{\alpha}$ in any compact set. This condition holds when h is one of $\underline{g}_k'$ or $\underline{g}_{ij}''$, for $i,j,k=1,\ldots,p$.

7) For $t,s=1,2,\ldots$, the absolute value of each element of the matrices $E(h(\underline{x}_t;\underline{\alpha})\underline{\varepsilon}_s^T)$, $E(h(\underline{x}_t;\underline{\alpha})\underline{c}_s^T)$ is less than $C\eta^{|t-s|}$, where C,η are defined in condition 5, and h in condition 6.

Then the MLE of $\underline{\alpha},\Omega$ are consistent, asymptotically normally distributed and efficient in the MP sense.

Proof:

The second partial derivatives are of the form 3.2 through 3.7, where a change of notation is made from Σ to Ω, and from $\underline{\varepsilon}_t$ to $\underline{u}_t=\underline{\varepsilon}_t-\underline{c}_t$. For future reference, $E(\underline{u}(n))=0$ and $E(\underline{u}(n)\underline{u}(n)^T)=\Omega(n)$, so that

3.24 $$E(\underline{u}_t\underline{u}_s^T) = \delta_{ts}\Omega \qquad \text{for } t,s=1,2,\ldots$$

The expected value of the term of the form 3.2 is

$$\omega^{hi}\,\omega^{jk} + \omega^{hj}\,\omega^{ik}$$

where ω^{ij} is the i,j'th element of Ω^{-1}. The convergence of the variance of this term to zero as $n \to \infty$ uniformly for $\underline{\alpha}$ in a compact set, also follows from 3.24. The expectations and variances of the terms corresponding to 3.3 and 3.4 behave similarly, by the same arguments.

Assumption 6 implies that the expectation of the term corresponding to 3.5 converges, as $n \to \infty$, uniformly for $\underline{\alpha},\Omega$ in a compact set. A representative term in the variance is

3.25 $$\operatorname{var}\Big[n^{-1}\sum_{t=1}^{n} \underline{g}_k'(\underline{x}_t;\underline{\alpha})^T(\Omega^{-1})_{.i}\,\underline{u}_t^T(\Omega^{-1})_{.j}\Big]$$

$$= n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} [(\Omega^{-1})_{i.}\,E(\underline{g}_k'(\underline{x}_t;\underline{\alpha})\,\underline{g}_k'(\underline{x}_s;\underline{\alpha})^T)(\Omega^{-1})_{.i}]*$$

$$[(\Omega^{-1})_{j.}\,E(\underline{u}_t\underline{u}_s^T)(\Omega^{-1})_{.j}]$$

$$+\ n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} [(\Omega^{-1})_{i.}\,E(\underline{g}_k'(\underline{x}_t;\underline{\alpha})\underline{u}_s^T)(\Omega^{-1})_{.j}]*$$

$$[(\Omega^{-1})_{i.}\,E(\underline{g}_k'(\underline{x}_s;\underline{\alpha})\underline{u}_t^T)(\Omega^{-1})_{.j}]$$

The first term of 3.25 goes to zero as $n \to \infty$ by assumptions 1,2, and 3, or, equivalently, 3.24. The second term of 3.25 goes to zero as $n \to \infty$ by assumption 7. For both terms, the convergence is uniform in a compact set of $\underline{\alpha},\Omega$. By these arguments each term in the variance

goes to zero as required. The term corresponding to 3.6 follows similarly.

The expected value of the term corresponding to 3.7 is

$$-n^{-1} \sum_{t=1}^{n} E[\underline{g}_{ih}''(\underline{x}_t;\underline{\alpha})^T \Omega^{-1} \underline{\varepsilon}_t]$$

$$+ n^{-1} \sum_{t=1}^{n} E[\underline{g}_i'(\underline{x}_t;\underline{\alpha})^T \Omega^{-1} \underline{g}_h'(\underline{x}_t;\underline{\alpha})]$$

Assumptions 4 and 6 imply the uniform convergence of these terms for $\underline{\alpha},\Omega$ in a compact set. Assumptions 1,2,3,5, and 7 imply that the variance goes to zero uniformly.

Applying theorem 1.3, we conclude that the MLE of $\underline{\alpha}$ and Ω have the required optimal properties.

3.7 Example of a Nonlinear Model under Various Assumptions on the Exogenous Variables and the Error Variables

An interesting example of a nonlinear model which is econometric in origin is proposed by Jöreskog and Goldberger [1975]. The model arises as follows. The scalar variable w_t is determined by

3.26 $$w_t = \underline{\alpha}^T \underline{x}_t + v_t \qquad \text{for } t=1,2,\ldots$$

where $\underline{x}_t$ and $\underline{\alpha}$ are Kx1 vectors and v_t is normally distributed with mean 0 and variance σ^2. The Lx1 vector $\underline{y}_t$ is determined as

3.27 $$\underline{y}_t = \underline{\beta} w_t + \underline{u}_t \qquad \text{for } t=1,2,\ldots$$

where $\underline{\beta}$ and $\underline{u}_t$ are Lxl vectors, and $\underline{u}_t$ is normally distributed with mean $\underline{0}$ and covariance matrix Θ. Combining 3.26 and 3.27, the authors obtain the model

$$\underline{y}_t = \underline{\beta}\ \underline{\alpha}^T \underline{x}_t + \underline{\varepsilon}_t \qquad \text{for } t=1,2,\ldots$$

where $\underline{\varepsilon}_t$ has the multivariate normal density of dimension L with mean $\underline{0}$ and covariance matrix $\Sigma = \sigma^2 \underline{\beta\beta}^T + \Theta$. The indeterminancy in the model parameters is removed by setting $\sigma^2 = 1$. The authors determine the MLE of the parameters when the exogenous variables are nonrandom and again when the exogenous variables are random and vary jointly with $\{\underline{y}_t\}$.

The theorems of this chapter may be applied to obtain optimal properties of the MLE under various assumptions.

1) If the exogenous variables are nonramdom and bounded above, $n^{-1} \sum_{t=1}^{n} \underline{x}_t \underline{x}_t^T$ converges, and the error terms are independently distributed, then the optimal properties of the MLE hold. This conclusion follows from theorem 3.1.

2) If the error variables form a multivariate process with mean $\underline{0}$ and covariance function $\rho^{|s|}\Sigma$, where $0 \leq \rho < 1$, the conditions involving $\underline{x}_t$ in 1 are satisfied, and if $n^{-1} \sum_{t=1}^{n} \underline{x}_t \underline{x}_{t-1}$ converges, the optimal properties of the MLE are guaranteed by theorem 3.2.

3) if $\underline{\varepsilon}_t$ and $\underline{x}_t$ are jointly distributed as multivariate normal deviates with mean $\underline{0}$ and covariance matrix

$$\begin{bmatrix} \Omega & \psi^T \\ \psi & \Phi \end{bmatrix}$$

for all t, where the covariance matrix is partitioned conformably with $\underline{\varepsilon}_t$ and $\underline{x}_t$, and $\underline{\varepsilon}_t, \underline{x}_t$ are serially independent, then the optimality properties of the MLE for $\underline{\alpha}, \underline{\beta}$, and $\Omega - \psi^T \Phi^{-1} \psi$ follow from theorem 3.5. A sequence of random variables $\{b_t\}$ is said to be serially independent if, for any t and any s, $t \neq s$, b_t and b_s are independently distributed.

CHAPTER IV: SYSTEMS OF EQUATIONS

4.1 Results from the Literature

The results for estimation of multivariate models presented in the previous chapters contribute to the theory of optimal estimation for simultaneous equation systems. The distinction to be made between general multivariate models and systems of simultaneous equations in structural form involves the estimation of coefficients of the endogenous variables in the latter model. Assumptions must be made to guarantee that theorem 1.4 applies, so that optimal estimators for the parameters of the structural model may be obtained from optimal estimators of the reduced form model. In this chapter, we will cite results concerning optimality properties of MLE from the literature, and modify theorems from the previous chapters to cover estimation of structural coefficients of systems.

Koopmans and Hood [1953] investigated the structural equation model

4.1 $$B\underline{y}_t + \Gamma\underline{z}_t = \underline{u}_t \qquad \text{for } t=1,2,\ldots$$

where B is an LxL nonsingular matrix of unknown parameters, Γ is an LxK matrix of unknown parameters, $\underline{y}_t$ and $\underline{u}_t$ are Lx1 vector variables, and $\underline{z}_t$ is a Kx1 vector of either lagged values of the endogenous variables or exogenous variables. The joint density of

the elements of $\underline{u}_t$ is multivariate normal with mean $\underline{0}$ and covariance matrix Σ. The disturbance terms $\underline{u}_t$ and $\underline{u}_s$ are independent for $t \neq s$ and are independent of all $\underline{z}_t$. The authors assume that all equations are identified; a detailed discussion of this concept is contained in section 4.2. Let $\underline{r}_t^T = (\underline{y}_t^T, \underline{z}_t^T)$. Let M_{ab} be the moment matrix of the variables $\underline{a}$ and $\underline{b}$ defined by

4.2 $$M_{ab} = n^{-1} \sum_{t=1}^{n} (\underline{a})_t (\underline{b})_t$$

Let W_{rr} be defined by

4.3 $$W_{rr} = M_{rr} - M_{rz} M_{zz}^{-1} M_{zr}$$

Then the MLE $\hat{A}$ of $A = [B, \Gamma]$ is shown to be that value of A which minimizes

4.4 $$V = \frac{\det(A M_{rr} A^T)}{\det(A W_{rr} A^T)}$$

The MLE of the covariance matrix Σ is given by

4.5 $$\hat{\Sigma} = \hat{A} M_{rr} \hat{A}^T$$

Assuming that M_{zz} approaches a nonsingular limit matrix in probability, the MLE are shown to be consistent and asymptotically normally distributed.

4.2 Applications of Theorems for Nonlinear Models to Linear Systems of Equations

We shall apply the results pertaining to nonlinear models in Chapter III to model 4.1 to prove optimality properties. The structural model 4.1 is transformed to its reduced form

$$\underline{y}_t = \Pi \underline{z}_t + \underline{\varepsilon}_t \qquad \text{for } t=1,2,\ldots \tag{4.6}$$

where $\Pi = -B^{-1}\Gamma$ and $\underline{\varepsilon}_t = B^{-1}\underline{u}_t$. The density of the vector $\underline{\varepsilon}_t$ is multivariate normal with mean $\underline{0}$ and covariance matrix $\Omega = B^{-1}\,\Sigma(B^{-1})^T$.

Following the standard usage, the equations of a system will be called identified if knowledge of the joint distribution of the observations $\underline{y}_1,\ldots,\underline{y}_n$ implies knowledge of all unknown parameters in the model. The definition may also be phrased as the existence of a one-to-one relationship between the parameters of the reduced equations and the parameters of the structural equations. Several authors give necessary and sufficient conditions for the identifiability of a single equation. For a particular equation of the system, let the variables be partitioned as

$$\underline{y}_t^T = (\underline{v}_t^T, \underline{w}_t^T)$$
$$\underline{z}_t^T = (\underline{x}_t^T, \underline{c}_t^T)$$
$$\underline{\varepsilon}_t^T = (\underline{\delta}_t^T, \underline{\xi}_t^T)$$

where $\underline{v}_t$ is the H-dimensioned subvector of $\underline{y}_t$ whose elements

have nonzero coefficients in the equation, $\underline{x}_t$ is the D-dimensional subvector of $\underline{z}_t$ whose elements have nonzero coefficients in the equation chosen, and $\underline{\varepsilon}_t$ is partitioned similarly to $\underline{y}_t$.

It is customary to assume, without loss of generality, that one element of $\underline{v}_t$ occurs with coefficient restricted to be one. Equivalently, it could be assumed that a quadratic form in the coefficients of $\underline{v}_t$ is restricted to be a constant.

The reduced form model may be written as

$$\begin{pmatrix} \underline{v}_t \\ \underline{w}_t \end{pmatrix} = \begin{bmatrix} \Pi_{vx} & \Pi_{vc} \\ \Pi_{wx} & \Pi_{wc} \end{bmatrix} \begin{pmatrix} \underline{x}_t \\ \underline{c}_t \end{pmatrix} + \begin{pmatrix} \underline{\delta}_t \\ \underline{\xi}_t \end{pmatrix} \qquad t=1,2,...$$

where the matrix Π is conformably partitioned. The equation chosen is identified if and only if the rank of Π_{vc} is H-1. A necessary condition for identification is that the number of linear restrictions on the parameters of the equation is equal to the number of equations in the system, less one. Consideration will be limited to those linear restrictions setting certain parameters of the structural equations equal to zero. If every equation of the system is identified, there exists a one-to-one map

$$\Psi: \ (\Pi,\Omega) \rightarrow (B,\Gamma,\Sigma)$$

This map clearly satisfies the smoothness conditions of theorem 1.4. Following are four theorems with various distributional assumptions on the exogenous and error variables.

Theorem 4.1

For the model 4.1, assume that the exogenous variables $\{\underline{z}_t,\ t=1,2,\ldots\}$ are nonrandom, and the error variables $\{\underline{u}_t,\ t=1,2,\ldots\}$ are independently distributed, each term having the multivariate normal distribution with expectation $\underline{0}$ and nonsingular covariance matrix Σ. Assume that B is nonsingular and that each equation is identified. Assume also that

1) $\{\underline{z}_t,\ t=1,2,\ldots\}$ is contained in a bounded set.

2) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \underline{z}_t \underline{z}_t^T$ exists.

Then the MLE of B, Γ, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Theorem 3.1 may be applied to model 4.6. In this case, $\underline{g}(\underline{z}_t;\Pi)$ is the $L\times 1$ vector with i'th element $(\Pi)_{i.}\underline{z}_t$. Then the first partial derivatives are

$$\partial \underline{g}(\underline{z}_t;\ \Pi)/\partial(\Pi)_{ij} = (0,\ldots,(\underline{z}_t)_j,\ldots,0)$$

where the only nonzero element is the i'th. Thus condition 1 of theorem 3.1 requires that

$$(\Omega^{-1})_{ik}\ n^{-1} \sum_{t=1}^{n} (\underline{z}_t)_j(\underline{z}_t)_m$$

converges, as $n \to \infty$, for $j,m=1,\ldots,K$ and $i,k=1,\ldots,L$. The second condition requires that $\{\underline{z}_t, t=1,2,\ldots\}$ be contained in a bounded set. The third condition is not restrictive since $\underline{g}$ is a multilinear function. Since the conditions of theorem 3.1 hold, the MLE of Π and Ω have the optimality properties claimed. By theorem 1.4, the MLE of B,Γ, and Σ also enjoy these optimality properties.

Theorem 4.2

Assume for model 4.1 that the exogenous variables are nonrandom and the error variables form a multivariate stationary process whose density is multivariate normal with mean $\underline{0}$ and covariance function $\rho^{|s|}\Sigma$ where ρ,Σ are unknown, $0\leq\rho<1$, and Σ positive definite. Assuming conditions 1 and 2 of theorem 4.1 and

$$3)\quad \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \underline{z}_t \underline{z}_{t-1}^T \quad \text{exists}$$

the MLE of B,Γ,Σ, and ρ are consistent, asymptotically normally distributed and efficient in the MP sense.

Proof:

Theorem 3.2. is applied to model 4.6. Therefore the MLE of Π,ρ,Ω have the optimality properties required. Applying theorem 1.4, the MLE of B,Γ,Σ, and ρ have the required properties.

Theorem 4.3

For model 4.1, assume the exogenous variables $\{\underline{z}_t, t=1,2,\ldots\}$ are random and identically distributed, independently of the error

terms, which are independently distributed, multivariate normal, with mean $\underline{0}$ and nonsingular covariance matrix Σ. Assume also that

1) All components of $|E(\underline{z}_t \underline{z}_s^T)|$ are bounded above, uniformly for $t,s=1,2,\ldots$

2) $|\mathrm{cov}((\underline{z}_t)_i(\underline{z}_t)_j, (\underline{z}_s)_i(\underline{z}_s)_j| \le C\eta^{|t-s|}$ for $i,j=1,\ldots,K$ where C is a nonnegative constant and $0 \le \eta < 1$.

Then the MLE of B, Γ, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

For the reduced model 4.6, condition 1 above implies condition 1 of theorem 3.3. Condition 2 of theorem 3.3 is satisfied since $\underline{g}(\underline{z}_t;\Pi)$ is multilinear. Condition 2 above implies condition 3 of theorem 3.3. Therefore the MLE of Π and Ω have the optimality properties required. Application of theorem 1.4 implies that B, Γ, and Σ have the optimality properties required.

<u>Theorem 4.4</u>

For model 4.1, assume that the exogenous variables $\{\underline{z}_t, t=1,\ldots\}$ are random and distributed independently of the error terms, which form a multivariate stationary process. The common distribution of each $\underline{u}_t$ is multivariate normal with mean $\underline{0}$ and nonsingular covariance matrix Σ. The covariance function

of the process is $\rho^{|s|}\Sigma$, where $0 \leq \rho < 1$. Assume also that

1) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E(\underline{z}_t \underline{z}_t^T)$, $\lim_{n\to\infty} n^{-1} \sum_{t=2}^{n} E(\underline{z}_t \underline{z}_{t-1}^T)$ exist.

2) The components of $|E(\underline{z}_t \underline{z}_s^T)|$ are bounded above uniformly for $t,s=1,2,\ldots$

3) $|\mathrm{cov}[(\underline{z}_u \underline{z}_u^T)_{ij}, (\underline{z}_v \underline{z}_v^T)_{hk}]| \leq C\eta^{|t-s|}$ for $i,j,h,k,=1,\ldots,K$, where $u=t,t-1$ and $v=s,s-1$, for $t,s=1,2,\ldots$, where $C > 0$ and $0\leq\eta<1$.

Then the MLE of B,Γ,ρ, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Theorem 3.4 is applied to model 4.6 to yield asymptotic properties for Π,ρ and Ω. Theorem 1.4 implies that the same properties hold for B,Γ,ρ, and Σ.

4.3 Systems with Lagged Dependent Variables

Hood and Koopmans [1953] require only that the variable $\underline{z}_t$ be independent of the error term $\underline{u}_t$. This condition is satisfied if lagged values of $\underline{y}_t$ are included among the elements of $\underline{z}_t$ and the disturbance terms are independent. However, theorems 4.1 through 4.4 do not allow lagged values of the dependent variables. To include this case, theorem 2.3 is applied to the reduced form of the model. The model in structural form will be denoted by

$$B_0\underline{y}_t = B_1\underline{y}_{t-1} + \ldots + B_G\underline{y}_{t-G} + \Gamma\underline{z}_t + \underline{u}_t \qquad t=1,2,\ldots \tag{4.7}$$

$B_0,\ldots,B_G$ are LxL matrices of unknown parameters; B_0 is nonsingular, Γ is an LxK matrix of unknown parameters. The vectors $\underline{y}_t$ and $\underline{u}_t$ are Lx1; the exogenous variable $\underline{z}_t$ is Kx1 dimensioned. The density of each error is multivariate normal with mean vector $\underline{0}$ and nonsingular covariance matrix Σ. The reduced form of 4.7 is

$$\underline{y}_t = C_1\underline{y}_{t-1} + \ldots + C_G\underline{y}_{t-G} + D\underline{z}_t + \underline{\varepsilon}_t \qquad t=1,2,\ldots \tag{4.8}$$

where $C_i = B_0^{-1}B_i$, for $i=1,\ldots,G$, $D=B_0^{-1}\Gamma$, and $\underline{\varepsilon}_t = B_0^{-1}\underline{u}_t$, for $t=1,2,\ldots$. The covariance matrix of $\underline{\varepsilon}_t$ is $\Omega = B_0^{-1}\Sigma(B_0^{-1})^T$. To estimate the structural parameters, each equation of the system must be identified. The transformation Ψ defined by

$$\Psi: (C_1,\ldots,C_G,D,\Omega) \rightarrow (B_0,\ldots,B_G,\Gamma,\Sigma)$$

is one-to-one and satisfies the assumptions of theorem 1.4.

Theorem 4.5

For the model 4.7, assume that the exogenous variables $\{\underline{z}_t, t=1,\ldots\}$ are nonrandom and the error terms $\{\underline{u}_t, t=1,\ldots\}$ are independently distributed, with multivariate normal density with mean $\underline{0}$ and nonsingular covariance matrix Σ. Assume B_0 is nonsingular, each equation of the system is identified, and the stability condition holds. Assume also that

1) $\{\underline{z}_t, t=1,\ldots\}$ is contained in a bounded set.

2) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n-c} \underline{z}_t \underline{z}_{t+c}^T$ exists, and the convergence is uniform in c, $c=0,1,\ldots$.

Then the MLE of $B_0, B_1, \ldots, B_G, \Gamma$, and Σ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Theorem 2.3 may be applied to model 4.8 to show that the MLE of the reduced form have the optimal properties required. Since the conditions of theorem 1.4 are satisfied by Ψ, the MLE of the structural parameters have the desired properties.

Optimality results can also be derived if the exogenous variables are random.

Theorem 4.6

For the model 4.7, assume that the exogenous variables are identically distributed, independently of the error sequence $\{\underline{u}_t, t=1,\ldots\}$. The error variables have the same distribution as specified in theorem 4.5. Assume that B_0 is nonsingular, each equation of the system is identified, and the stability condition holds. Assume also that conditions 1,2, and 3 of theorem 2.2 hold when x is replaced by z. Then the MLE of $B_0, B_1, \ldots, B_G, \Gamma$, and Σ are consistent, asymptotically normally distributed and efficient in the MP sense.

Proof:

The assumptions of theorems 2.4 and 1.4 are satisfied; the result follows.

4.4 Optimality Properties when Independence between the Error Variables and the Exogenous Variables is not Assumed

Dependence between the variables $\{\underline{z}_t, t=1,\ldots\}$ and $\{\underline{u}_t, t=1,\ldots\}$ may also be allowed in cases other than those when components of $\underline{z}_t$ are lagged values of $\underline{y}_t$. To obtain optimal properties for the MLE in these more general cases, theorem 3.5 will be adapted for systems of equations.

Theorem 4.7

For model 4.6, assume that the joint distribution of $\underline{\varepsilon}(n)$ and $\underline{z}(n)$ is multivariate normal with mean zero and covariance matrix

$$\begin{bmatrix} \Sigma_{\varepsilon}(n) & \Sigma_{\varepsilon z}^{T}(n) \\ \Sigma_{\varepsilon z}(n) & \Sigma_{z}(n) \end{bmatrix}$$

The conditional mean of $\underline{\varepsilon}(n)$ given $\underline{z}(n)$ is defined to be $\underline{c}(n) = \Sigma_{\varepsilon z}(n)\,\Sigma_{z}(n)^{-1}\,\underline{z}(n)$.

Assume the following:

1) $\mathrm{var}\,(\underline{\varepsilon}_t - \underline{c}_t) = \mathrm{var}(\underline{\varepsilon}_s - \underline{c}_s)$ for all $t,s=1,\ldots$

2) $\mathrm{cov}\,(\underline{\varepsilon}_t, \underline{\varepsilon}_s) = \mathrm{cov}(\underline{c}_t, \underline{c}_s)$ for all $s \neq t$.

3) The elements on the diagonal of $\Sigma_{z}(n)$ are bounded above in absolute value, uniformly for all n.

4) $n^{-1} \sum_{t=1}^{n} E(\underline{z}_t \underline{z}_t^T)$ converges as $n \to \infty$.

5) $|\mathrm{cov}((\underline{z}_t)_i (\underline{z}_t)_j,\ (\underline{z}_s)_i (\underline{z}_s)_j)| \leq C\eta^{|t-s|}$,

for all $i,j=1,\ldots,K$,

for all t,s, where $C > 0$ and $0 \leq \eta < 1$.

6) $n^{-1} \sum_{t=1}^{n} E(\underline{z}_t \underline{\varepsilon}_t^T)$ converges as $n \to \infty$.

7) $|E((\underline{z}_t)_j (\underline{\varepsilon}_s)_k)| \leq C\eta^{|t-s|}$, for $j=1,\ldots,K$,

for $k=1,\ldots,L$,

for all $t, s=1,\ldots$.

Assuming conditions 1 through 7, when all equations are identified, the MLE of B,Γ, and $\Omega(n) = \Sigma_{\varepsilon}(n) - \Sigma_{\varepsilon z}(n)\, \Sigma_z(n)^{-1}\, \Sigma_z\, (n)$ are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

Theorem 3.5 may be applied to the reduced model 4.6. Assumptions 1 through 7 above imply assumptions 1 through 7 of theorem 3.5. Optimal properties of the structural parameters may be inferred from theorem 1.4 from the optimal properties of the reduced form parameters.

4.5 Estimation of the Parameters of a Single Equation or a Subsystem

Several authors have considered the problem of estimating the parameters of a subset or a single equation of a system of equations. Maximum likelihood methods for this purpose are called limited information methods, becuase they do not make use of restrictions

on parameters which do not occur in the subsystem. In contrast, maximum likelihood estimation of all parameters of an identified system is called a full information method. Previous research on asymptotic properties for the limited information method under very general distributional assumptions will be cited. Theorems from the previous sections will be adapted to subsystems.

Anderson and Rubin [1949], [1950] obtain the MLE for the parameters of a single identified equation under the assumption that the error terms for the complete system are normally distributed, with mean zero and covariance matrix Σ. Asymptotic properties for this estimator are then proven under more general conditions. The equation under consideration is denoted as

$$\underline{\beta}^T \underline{y}_t + \underline{\gamma}^T \underline{z}_t = \zeta_t \qquad t=1,2,\ldots$$

The variables $\underline{y}_t$ and $\underline{z}_t$ are partitioned as before into subvectors whose elements do appear in the chosen equation with nonzero coefficients, and those whose elements do not. The authors define $\underline{s}_t$ to be a linear transform of $\underline{c}_t$ such that $M_{xs} = 0$, where M_{xs} is defined in 4.2. The vector $\underline{b}$ is chosen to satisfy

$$(M_{vs} M_{ss}^{-1} M_{sv} - \nu W_{vv})\underline{b} = \underline{0}$$

where W_{vv} is defined in 4.3 and ν is the smallest root of the determinental equation

$$|M_{vs}M_{ss}^{-1}M_{sv} - \nu W_{vv}| = 0$$

The matrix Φ is the matrix of the quadratic form normalization for $\underline{\beta}$. Then the MLE for $\underline{\beta}$ is given as

$$\hat{\underline{\beta}} = \underline{b}/(\underline{b}^T\Phi\underline{b})^{1/2}$$

and the MLE for $\underline{\gamma}$ is

$$\hat{\underline{\gamma}} = -\hat{\underline{\beta}}\, M_{vx}M_{xx}^{-1}$$

Assuming that the chosen equation is identified and

AR1) $n^{-1}\sum_{t=1}^{n} \underline{z}_t\underline{z}_t^T$ converges stochastically to a nonsingular matrix R.

AR2) $n^{-1}\sum_{t=1}^{n} \underline{\delta}_t\underline{z}_t^T$ converges stochastically to 0.

AR3) The ratio of the largest to the smallest eigenvalues of

$$W_{vv}^{*} = M_{vv} - (M_{vx}\ M_{vs})\begin{bmatrix} M_{xx} & M_{xs} \\ M_{sx} & M_{ss} \end{bmatrix}^{-1}\begin{pmatrix} M_{xv} \\ M_{sv} \end{pmatrix}$$

is bounded above in probability.

The authors prove that $\hat{\underline{\beta}}$ and $\hat{\underline{\gamma}}$ are consistent. Under further assumptions which specify that

AR4) The sequence $\{\underline{z}_t\}$ is bounded and may include lagged endogenous variables, provided they are bounded almost surely.

AR5) For some $\lambda > 0$ and for some M, $E(|(\underline{\delta}_t)_i|^{4+\lambda}) < M$

AR6) For each i,j,k, and m,

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E((\underline{\delta}_t)_i(\underline{\delta}_t)_j(\underline{z}_t)_k(\underline{z}_t)_m) \text{ exists,}$$

$n^{1/2}(\hat{\underline{\beta}} - \underline{\beta})$, $n^{1/2}(\hat{\underline{\gamma}} - \underline{\gamma})$ are asymptotically normally distributed. No claims of efficiency are made; the covariance matrix Σ is not estimated. If Φ, the matrix of the quadratic form normalization, is not constant, but is a function of the covariance matrix of the error terms, then the following additional assumptions are required to conclude asymptotic normality:

AR7) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E((\underline{\delta}_t)_i(\underline{\delta}_t)_j(\underline{\delta}_t)_k(\underline{z}_t)_m)$ exists.

AR8) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E((\underline{\delta}_t)_i(\underline{\delta}_t)_j)$ exists.

AR9) $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E((\underline{\delta}_t)_i(\underline{\delta}_t)_j(\underline{\delta}_t)_k(\underline{\delta}_t)_m)$ exists.

If it is assumed that the disturbance vectors are distributed independently and normally with mean vector zero and covariance matrix Σ, and $\Phi = \Sigma$, then assumptions AR1, AR2, AR4, and AR5 are sufficient for the asymptotic normality of $\hat{\underline{\beta}}$ and $\hat{\underline{\gamma}}$.

Rubin and Chernoff [1953] also prove asymptotic properties for the MLE of a subsystem, computed under the assumptions of normality

and independence of the error terms. For the model 4.1, the variables are partitioned as before on the basis of whether or not they appear in the subsystem. The submatrices of B, Γ appearing in the subsystem will be denoted B_1, Γ_1. The following assumptions are required.

RC1) $n^{-1} \sum_{t=1}^{n} \underline{\delta}_t \underline{\delta}_t^T$ converges stochastically to a nonsingular matrix Σ_1.

RC2) $n^{-1} \sum_{t=1}^{n} \underline{z}_t \underline{z}_t^T$ converges stochastically to a nonsingular matrix.

RC3) $n^{-1} \sum_{t=1}^{n} \underline{\delta}_t \underline{z}_t^T$ converges stochastically to zero.

RC4) $n^{-1} \sum_{t=1}^{n} E(\underline{y}_t / \underline{y}_1, \ldots, \underline{y}_{t-1}, \underline{z}_1, \ldots, \underline{z}_t) \underline{z}_t^T$ converges stochastically.

RC5) $n^{-1} \sum_{t=1}^{n} [\underline{y}_t - E(\underline{y}_t / \underline{y}_1, \ldots, \underline{y}_{t-1}, \underline{z}_1, \ldots, \underline{z}_t)] \; \underline{z}_t^T$ converges stochastically to zero.

RC6) The matrix equation $[B_1, \Gamma_1]M = 0$ defines B_1, Γ_1 as a single valued differentiable function of M, for all M for which the matrix equation has a solution which satisfies the a priori restrictions.

Assuming conditions RC1 through RC6, the estimators of B_1, Γ_1, and Σ_1 computed by the ML method, assuming that the errors are independent and normally distributed, are consistent. To conclude the asymptotic normality of $\hat{B}_1$ and $\hat{\Gamma}_1$, two further assumptions are required.

RC7) The elements of $n^{-(1/2)} \sum_{t=1}^{n} \underline{\delta}_t \underline{z}_t^T$ are asymptotically normally distributed.

RC8) Let $C_{ik,jm}$ represent the asymptotic covariance of $n^{-(1/2)} \sum_{t=1}^{n} (\underline{\delta}_t)_i (\underline{z}_t)_k$ and $n^{-(1/2)} \sum_{t=1}^{n} (\underline{\delta}_t)_j (\underline{z}_t)_m$. Then $C_{ik,jm} - [n^{-1} \sum_{t=1}^{n} (\underline{\delta}_t)_i (\underline{\delta}_t)_j][n^{-1} \sum_{t=1}^{n} (\underline{z}_t)_k (\underline{z}_t)_m]$ converges stochastically to zero, for all i,j,k,m.

We apply our theorems on the estimation of the parameters of a complete system to the problem of estimating a subsystem. We will suppose that the subsystem consists of the first L_1 equations of the complete system 4.1, where $1 \leq L_1 \leq L-1$. We will assume that each of the L_1 equations is identified. Let $\underline{y}_t^*$ be composed of the elements of $\underline{y}_t$ which are present in the subsystem. Let Π^* be that submatrix of Π which consists of the first L_1 rows. Let $\underline{\varepsilon}_t^*$ be the subvector of $\underline{\varepsilon}_t$ which consists of the first L_1 elements, and Ω^* be the submatrix of Ω consisting of the first L_1 rows and columns. Ω^* is assumed to be invertible. If the error terms $\underline{\varepsilon}_t^*$ are independently distributed with mean vector zero, and the exogenous variables are either nonrandom or distributed independently of $\{\underline{\varepsilon}_t^*, t=1,\ldots\}$, then the logarithm of the likelihood function of $\underline{y}_1^*,\ldots,\underline{y}_n^*$ is the sum of a term which is not a function of the unknown parameters and

$$-(n/2)\log|\Omega^*| - (1/2) \sum_{t=1}^{n} (\underline{y}_t^* - \Pi^*\underline{z}_t)^T (\Omega^*)^{-1} (\underline{y}_t^* - \Pi^*\underline{z}_t)$$

By theorem 1.4, if the MLE of Π^*, Ω^* are consistent and asymptotically normally distributed, and efficient in the MP sense, then the MLE of the structural parameters of the subsystem also enjoy these properties. A similar argument holds when the $\{\underline{\varepsilon}_t^*, t=1,\ldots\}$ form a multivariate stationary process.

In summary, theorems 4.1 through 4.4 hold for subsystems when assumptions on $\{\underline{\varepsilon}_t, t=1,\ldots\}$ are assumed to hold only for the subvector $\{\underline{\varepsilon}_t^*, t=1,\ldots\}$, and when only the submatrix Ω^* is assumed to be invertible. The distributional assumptions on the exogenous variables remain unchanged. The same statement is true of the adaptations of theorems 4.5 and 4.6 to subsystems, if among the components of $\underline{z}_t$ are included only lagged values of $\underline{y}_t^*$ and not of the full vector $\underline{y}_t$. The adaptation of theorem 4.7 follows similarly.

4.6 Singular Systems

A singular system is defined to be one in which the covariance matrix of the disturbance terms is singular. In this case, the density of the error terms does not exist, when the distribution is assumed to be normal. Singular systems result when a linear combination of the components of the endogenous variables is constrained to be a constant. As an example, consider the model

$$\underline{y}_t = \underline{b} + \Gamma \underline{x}_t + \underline{\varepsilon}_t \qquad \text{for } t=1,\ldots$$

where $\underline{j}^T \underline{y}_t = 1$ for all t, with $\underline{j}$ being the Lx1 vector whose

components are all unity, $\underline{b}$ is Lx1 dimensioned, Γ is LxK dimensioned, $\underline{x}_t$ is Kx1 dimensioned, and $\underline{\varepsilon}_t$ is Lx1 dimensioned. Models of this type are considered in Barten [1969], where the error vectors are assumed to be independently normally distributed with mean vector $\underline{0}$ and covariance matrix Σ. Berndt and Savin [1975] consider this model with errors generated from a stationary vector process satisfying

$$\underline{\varepsilon}_t = R\underline{\varepsilon}_{t-1} + \underline{u}_t \qquad \text{for } t=2,3,\ldots$$

where the $\{\underline{u}_t\}$ are independently, identically distributed, multivariate normal with mean $\underline{0}$ and covariance matrix Σ. The constraint that $\underline{j}^T\underline{y}_t = 1$ implies

$$4.9 \qquad \underline{j}^T\underline{b}=1, \qquad \underline{j}^T\Gamma=\underline{0}, \qquad \underline{j}^T\underline{\varepsilon}_t=0.$$

The last equations imply that the covariance matrix is singular, for both error models described above. Barten proposes that one equation of the system be eliminated from the model, removing the singularity from the covariance matrix. Denoting by $\underline{a}_t^*$ the first L-1 elements of the vector $\underline{a}_t$, and by Γ^* the first L-1 rows of Γ, the model with independent error terms is

$$\underline{y}_t^* = \underline{b}^* + \Gamma^*\underline{x}_t + \underline{\varepsilon}_t^* \qquad \text{for } t=1,\ldots$$

where the covariance of $\underline{\varepsilon}_t^*$ is Σ^*, the L-1xL-1 nonsingular matrix

formed from Σ by removing the last row and last column. To obtain the MLE of $\underline{b},\Gamma$ from $\underline{b}^*,\Gamma^*$, the restrictions 4.9 are used. Barten has derived the transformation from Σ^* to Σ. This transformation is

$$\begin{bmatrix} I_{L-1} & \begin{matrix}-1\\ \vdots \\ -1\end{matrix} \\ -1 \cdots\cdots -1 & -1 \end{bmatrix} \begin{bmatrix} \Sigma^* & \begin{matrix}0\\ \vdots \\ 0\end{matrix} \\ 0 \cdots\cdots 0 & L^{-1} \end{bmatrix} \begin{bmatrix} I_{L-1} & \begin{matrix}-1\\ \vdots \\ -1\end{matrix} \\ -1 \cdots\cdots -1 & -1 \end{bmatrix} = \Sigma + L^{-1}\underline{j}\underline{j}^T$$

It is obvious that the map

$$\psi: \ (\underline{b}^*, \Gamma^*, \Sigma^*) \to (\underline{b}, \Gamma, \Sigma)$$

is bijective and satisfies the smoothness conditions of theorem 1.4. Thus the theorems of sections 4.2 and 4.4 may be applied to the truncated form of a singular system to obtain optimal properties of the MLE for the full model.

4.7 Seemingly Unrelated Equations

A system of seemingly unrelated equations has certain simplifying properties. The model is

4.10 $$(\underline{y}_t)_i = \underline{z}_{ti}^T \underline{\beta}_i + (\underline{u}_t)_i \qquad i=1,\ldots,L;\ t=1,\ldots$$

where $\underline{z}_{ti}$ is a K_ix1 vector of exogenous variables, and $\underline{\beta}_i$ is the K_ix1 vector of unknown parameters. Let

$$\underline{z}_t^T = (\underline{z}_{t1}^T, \ldots, \underline{z}_{tL}^T) \quad \text{and} \quad K = K_1 + \ldots + K_L.$$

The model 4.10 may be written in multivariate form as

$$\underline{y}_t = \begin{bmatrix} \underline{z}_{t1}^T & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \underline{z}_{tL}^T \end{bmatrix} \begin{bmatrix} \underline{\beta}_1 \\ \vdots \\ \underline{\beta}_L \end{bmatrix} + \underline{u}_t \qquad t=1,2,\ldots$$

The term "seemingly unrelated" refers to the fact that if the covariance matrix of $\underline{u}_t$ were diagonal for each t, the MLE of $\underline{\beta}_i$ would be a function of $(\underline{y}_t)_i$, $\underline{z}_{ti}$, $t=1,\ldots,n$, alone. The representation of the model as a system would not produce more efficient (in any sense) estimators. It will be assumed in this section that the covariance matrix of $\underline{u}_t$ is not diagonal, for $t=1,2,\ldots$. Assuming the appropriate conditions on the error terms and the exogenous variables, theorems 3.1 through 3.5 may be applied.

Recent interest has been expressed in the efficient estimation of the parameters of seemingly unrelated equations when the error terms satisfy an autoregressive process. Parks [1967] and Kmenta and Gilbert [1970] consider the case where each component of $\underline{u}_t$ satisfies a first order autoregressive process

4.11 $$(\underline{u}_t)_j = \rho_j(\underline{u}_{t-1})_j + (\underline{\varepsilon}_t)_j \qquad j=1,\ldots,L;\ t=2,3,\ldots$$

$$(\underline{u}_1)_j = (1-\rho_j)^{-(1/2)}(\underline{\varepsilon}_1)_j \qquad j=1,\ldots,L$$

The variables $\underline{\varepsilon}_t$ are independently distributed, with mean vector

zero and nonsingular covariance matrix. The stability condition is assumed to hold; i.e., $|\rho_i| < 1$, for $i=1,\ldots,L$. Parks proposes a three-step estimation technique which is consistent and asymptotically efficient in the sense of attaining the Cramér-Rao lower bound on the covariance matrix. For this model, Kmenta and Gilbert compare small sample efficiency for several alternative estimators, one of which is Parks' estimator.

Guilkey and Schmidt [1973] treat the case of vector autoregressive models

4.11 $$\underline{u}_t = R\underline{u}_{t-1} + \underline{\varepsilon}_t \qquad t=2,3,\ldots$$

where $\{\underline{\varepsilon}_t, t=1,\ldots\}$ are distributed independently with mean $\underline{0}$ and nonsingular covariance matrix. The stability condition is assumed to hold; the L roots of the determinental equation

4.12 $$|\rho - R| = 0$$

are less than one in absolute value. The authors propose a six step procedure, similar to the procedure of Parks, which produces a consistent, asymptotically efficient (in the sense of attaining the Cramér-Rao lower bound on the covariance matrix) estimator of $\underline{\beta}$.

It should be noted that these results apply to autoregressive error structures of order one. Using theorems 2.3, 2.4, and 1.4, we can obtain optimal properties for the MLE for general autoregressive error structures. Theorem 4.8 below may be applied to a system

of seemingly unrelated equations with autoregressive error structure, or to a system of identified linear equations with autoregressive error structure. Hendry [1971] obtained the MLE and their asymptotic covariance matrix for a model of the latter type.

Theorem 4.8

For model 4.1, where each equation is assumed to be identified, or model 4.10, assume the error variables $\{\underline{u}_t,\ t=1,\ldots\}$ satisfy

$$\underline{u}_t = R_1\underline{u}_{t-1} + \ldots + R_H\underline{u}_{t-H} + \underline{\varepsilon}_t \qquad t=1,\ldots$$

where the $\{\underline{\varepsilon}_t,\ t=1,\ldots\}$ are assumed to be independently normal with mean vector zero and covariance matrix Σ. The roots of the determinental equation

$$|\rho^H - R_1\rho^{H-1} - \ldots - R_H| = 0$$

are assumed to be less than one in absolute value. If the exogenous variables are nonrandom and

1) $\{\underline{z}_t,\ t=1,\ldots\}$ is contained in a bounded set,
2) $\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \underline{z}_t\underline{z}_{t+c}^T$ exists for all $c = 0,1,\ldots$;

the convergence is uniform in c,

then the MLE of B, Γ, $R_1, \ldots, R_H$, and Σ for model 4.1, or $\underline{\beta}$, $R_1, \ldots, R_H$, and Σ for model 4.10 are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

By the method of section 2.5, and by theorem 2.3, the result follows.

Theorem 4.9

For the same models and error structure as in theorem 4.8, assume that the exogenous variables are identically distributed, independently of $\{\underline{\varepsilon}_t, t=1,\ldots\}$. Assume also conditions 1,2, and 3 of theorem 2.2, where x is replaced by z. Then the MLE of the parameters of the model are consistent, asymptotically normally distributed, and efficient in the MP sense.

Proof:

By the method of section 2.5 and by theorem 2.4, the MLE have the required properties.

CHAPTER V: PROBIT AND LOGIT MODELS: TOPICS FOR FUTURE RESEARCH

5.1 Introduction

The models covered in this chapter have recently appeared in the literature. In order to explain complicated behavior of the variables in these models, the authors make complex distributional assumptions for the observations. They suggest maximum likelihood as the method of estimation, but do not conjecture or prove optimality properties. We will examine the nature of assumptions required in order to invoke theorems 1.1, 1.2. and 1.3. Sections 5.2, 5.3, and 5.4 deal with variations of the probit model. Section 5.5 treats a simultaneous logit model.

5.2 A One-Limit Probit Model

Tobin [1958] deals with the one-limit probit model, where the underlying distribution is assumed to be normal. Amemiya and Boskin [1974] use the lognormal as the underlying distribution for their one-limit probit model, to account for the nonnegativity and skewness of their dependent variable. The authors postulate that the independent observations $\{y_t, t=1,\ldots\}$ satisfy

$$y_t = z_t \quad \text{if} \quad z_t < L;$$

$$y_t = L \quad \text{if} \quad z_t \geq L,$$

for $t = 1,2,\ldots,$ where the distribution of z_t is independent log-normal, with expectation $\underline{\theta}^T \underline{x}_t$, where $\underline{\theta}$ is a Kx1 vector of unknown parameters and $\underline{x}_t$ is a Kx1 vector of exogenous variables, for all t. The variance of z_t, for all t, is proportional to the square of the mean, $C^2(\underline{\theta}^T \underline{x}_t)^2$. The density of z_t, for all t, is given by

$$g_t(z_t) = \frac{1}{(2\pi)^{1/2}\sigma\, z_t} \exp\left(\frac{-[\log(z_t)-\log(\underline{\theta}^T \underline{x}_t) + \sigma^2/2]^2}{2\sigma^2}\right)$$

where $\sigma^2 = \log(1+C^2)$. Let F represent the normal c.d.f. where the mean is zero and the variance is σ^2. Let f be the density of F. Let

$$v_t = \log(y_t) - \log(\underline{\theta}^T \underline{x}_t) + \sigma^2/2$$

$$u_t = -\log(L) + \log(\underline{\theta}^T \underline{x}_t) - \sigma^2/2.$$

Finally, let $S_1 = \{t\colon\ z_t \geq L\}$, and $S_2 = \{t\colon\ z_t < L\}$. Let T_2 be the number of elements in S_2. The log likelihood function of the observations, except for a constant with respect to the unknown parameters, is

$$\log L_n(\underline{\theta}, \sigma^2) = \sum_{S_1} \log F(u_t) - (1/2) \sum_{S_2} \log(\sigma^2) - (2\sigma^2)^{-1} \sum_{S_2} v_t^2.$$

Let F_t, f_t denote F and f, respectively, evaluated at u_t. The authors compute the second derivatives of the log likelihood function.

5.1 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial\underline{\theta}\partial\underline{\theta}^T} = (n\sigma^2)^{-1}\sum_{S_1}\frac{u_t f_t F_t + \sigma^2 f_t(f_t+F_t)}{F_t^2(\underline{\theta}^T\underline{x}_t)^2}\underline{x}_t\underline{x}_t^T$$

$$+ (n\sigma^2)^{-1}\sum_{S_1}\frac{1+v_t}{(\underline{\theta}^T\underline{x}_t)^2}\underline{x}_t\underline{x}_t^T$$

5.2 $$\frac{-n^{-1}\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial\sigma^2\partial\underline{\theta}} = -(2\sigma^4 n)^{-1}\sum_{S_1}\left\{\frac{(u_t+\sigma^2)(\sigma^2 f_t^2+u_t f_t F_t)-\sigma^2 f_t F_t}{F_t^2(\underline{\theta}^T\underline{x}_t)}\right\}\underline{x}_t$$

$$-(2\sigma^4 n)^{-1}\sum_{S_2}\left\{\frac{\sigma^2-2v_t}{\underline{\theta}^T\underline{x}_t}\right\}\underline{x}_t$$

5.3 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial^2(\sigma^2)} =$$

$$(-4\sigma^4 n)^{-1}\sum_{S_1}\left\{\frac{f_t}{F_t^{\,2}}\right\}\left\{(u_t+\sigma^2)^2\left\{\frac{u_t F_t}{\sigma^2}+f_t\right\}-(3u_t+2\sigma^2)F_t\right\}$$

$$-(T_2/(2\sigma^4 n)) + (T_2/(4\sigma^2 n))$$

$$-\sigma^{-4}n^{-1}\sum_{S_2} v_t + \sigma^{-6}n^{-1}\sum_{S_2} v_t^2.$$

To invoke theorems 1.1, 1.2, and 1.3, the expectations of terms 5.1, 5.2, and 5.3 must converge as $n \to \infty$, uniformly for $\underline{\theta}$, σ^2 in any compact set. Also, the variances must converge to zero as $n \to \infty$,

uniformly in the same sense. If the exogenous variables are nonrandom, the randomness of the first term of 5.1 is due to the fact that the summation is over a random set S_1. Let $I(E)$ denote the indicator function which assumes the value 1 if the event E occurs, and zero otherwise. The summation over the random set S_1 can be written as the summation, over all t, of $I(z_t \geq L)$ times the t'th term of the summand. Since

$$E(I(z_t \geq L)) = F_t$$

the expectation of the first term of 5.1 is

$$(n\sigma^2)^{-1} \sum_{t=1}^{n} \left\{ \frac{u_t f_t F_t + \sigma^2 f_t (f_t + F_t)}{F_t (\underline{\theta}^T \underline{x}_t)^2} \right\} \underline{x}_t \underline{x}_t^T .$$

We must assume that this term converges uniformly as $n \to \infty$, for all $\underline{\theta}$, σ^2 in a compact set. The complexity of the summation makes it difficult to simplify the assumption further. The expectation of the second term of 5.1 is

$$(n\sigma^2)^{-1} \sum_{t=1}^{n} \left\{ \frac{(1-\log(\underline{\theta}^T \underline{x}_t) + \sigma^2/2)(1-F_t) + \int_{-\infty}^{L} \log(z) g_t(z) dz}{(\underline{\theta}^T \underline{x}_t)^2} \right\} \underline{x}_t \underline{x}_t^T .$$

This term also must be assumed to converge uniformly as $n \to \infty$, for all $\underline{\theta}$, σ^2 in a compact set. The expectation of 5.2 is

$$- (2\sigma^4 n)^{-1} \sum_{t=1}^{n} \left\{ \frac{(u_t + \sigma^2)(\sigma^2 f_t^2 + u_t f_t F_t) - \sigma^2 f_t F_t}{F_t(\underline{\theta}^T \underline{x}_t)} \right\} \underline{x}_t$$

$$- (2\sigma^4 n)^{-1} \sum_{t=1}^{n} \left\{ \frac{-2\log(\underline{\theta}^T \underline{x}_t)(1-F_t) - 2 \int_{-\infty}^{L} \log(z) g_t(z) dz}{\underline{\theta}^T \underline{x}_t} \right\} x_t .$$

The expectation of 5.3 is

$$(4\sigma^4 n)^{-1} \sum_{t=1}^{n} f_t [(u_t + \sigma^2)^2 (f_t F_t + u_t/\sigma^2) - (3u_t + 2\sigma^2)]$$

$$+ (4\sigma^2)^{-1} - (2\sigma^4)^{-1} + [(2\sigma^4)^{-1} - (4\sigma^2)^{-1}] n^{-1} \sum_{t=1}^{n} F_t .$$

The two terms above must be assumed to converge as before. Since the $\{e_t, t=1,\ldots\}$ are independently distributed, assuming that $\{x_t, t=1,\ldots\}$ is contained in a bounded set is sufficient to conclude that the variances of 5.1, 5.2, and 5.3 converge to zero, as $n \to \infty$, uniformly for $\underline{\theta}, \sigma^2$ in a compact set.

5.3 A Two-Limit Probit Model

Rosett and Nelson [1975] consider a model in which the endogenous variable is restricted by an upper as well as a lower limit, but is continuous between the limits. The behavior of the variable between the limits is explained by a linear function of exogenous variables; we will relax this specification to include nonlinear functions as well.

Let $\{z_t, t=1,\ldots\}$ be independently distributed normal variates with expectation $g(\underline{x}_t; \underline{\theta})$ and variance $\sigma^2 > 0$. The function g is not necessarily linear but must satisfy certain differentiability conditions to be determined below. Let F and f be as defined in section 5.2. Let $\{L_{1t}, t=1,\ldots\}$ and $\{L_{2t}, t=1,\ldots\}$ be sequences of real numbers with $L_{1t} < L_{2t}$ for all t. The behavior of the dependent variables $\{y_t, t=1,\ldots\}$ is given by

$$y_t = L_{1t} \quad \text{if} \quad z_t < L_{1t}$$

$$y_t = L_{2t} \quad \text{if} \quad z_t > L_{2t}$$

$$y_t = z_t \quad \text{if} \quad L_{1t} \leq z_t \leq L_{2t}.$$

Let

$$S_1 = \{t:\ z_t < L_{1t}\};\ S_2 = \{t:\ z_t > L_{2t}\};\ S_3 = \{t:\ L_{1t} \leq z_t \leq L_{2t}\}.$$

Let $u_{1t} = g(\underline{x}_t; \underline{\theta}) - L_{1t}$ and $u_{2t} = g(\underline{x}_t; \underline{\theta}) - L_{2t}$. Then the logarithm of the likelihood function is

$$\begin{aligned}\log L_n(\underline{\theta}, \sigma^2) = & \sum_{S_1} \log(1-F(u_{1t})) \\ & + \sum_{S_2} \log(F(u_{2t})) \\ & + \sum_{S_3} [-\log(\sigma) - (2\sigma^2)^{-1}(y_t - g(\underline{x}_t; \underline{\theta}))^2].\end{aligned}$$

As in Chapter III, we denote $g_i'(\underline{x}_t; \underline{\theta}) = \partial g(\underline{x}_t; \underline{\theta})/\partial\theta_i$ and $g_{ij}''(\underline{x}_t; \underline{\theta}) = \partial^2 g(\underline{x}_t; \underline{\theta})/\partial\theta_i\partial\theta_j$. The second partial derivatives of the log likelihood function are

5.4 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial\theta_i\partial\theta_j} =$$

$$n^{-1}\sum_{S_1}\left\{g_i'(\underline{x}_t;\underline{\theta})g_j'(\underline{x}_t;\underline{\theta})\left(\frac{f^2(u_{1t})}{(1-F(u_{1t}))^2}-\frac{u_{1t}f(u_{1t})}{\sigma^2(1-F(u_{1t}))}\right)\right.$$
$$\left.+\frac{f(u_{1t})}{1-F(u_{1t})}g_{ij}''(\underline{x}_t;\underline{\theta})\right\}$$

$$+n^{-1}\sum_{S_2}\left\{g_i'(\underline{x}_t;\underline{\theta})g_j'(\underline{x}_t;\underline{\theta})\left(\frac{f^2(u_{2t})}{F^2(u_{2t})}+\frac{u_{2t}f(u_{2t})}{\sigma^2F(u_{2t})}\right)-\frac{f(u_{2t})}{F(u_{2t})}g_{ij}''(\underline{x}_t;\underline{\theta})\right\}$$

$$+(n\sigma^2)^{-1}\sum_{S_3}\left\{g_i'(\underline{x}_t;\underline{\theta})g_j'(\underline{x}_t;\underline{\theta})-(y_t-g(\underline{x}_t;\underline{\theta}))g_{ij}''(\underline{x}_t;\underline{\theta})\right\}$$

5.5 $$-n^{-1}\frac{\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial\theta_i\partial\sigma} =$$

$$-n^{-1}\sum_{S_1}g_i'(\underline{x}_t;\underline{\theta})\left\{\frac{u_{1t}f^2(u_{1t})}{\sigma(1-F(u_{1t}))^2}-\frac{f(u_{1t})[(u_{1t}/\sigma)^2-1]}{\sigma(1-F(u_{1t}))}\right\}$$

$$-n^{-1}\sum_{S_2}g_i'(\underline{x}_t;\underline{\theta})\left\{\frac{u_{2t}f^2(u_{2t})}{\sigma F^2(u_{2t})}+\frac{f(u_{2t})[(u_{2t}/\sigma)^2-1]}{\sigma F(u_{2t})}\right\}$$

$$-2(\sigma^3 n)^{-1}\sum_{S_3}g_i'(\underline{x}_t;\underline{\theta})(y_t-g(\underline{x}_t;\underline{\theta})).$$

5.6 $$- n^{-1} \frac{\partial^2 \log L_n(\underline{\theta},\sigma^2)}{\partial^2 \sigma} =$$

$$- n^{-1} \sum_{S_1} \left\{ \frac{u_{1t}^2 f^2(u_{1t})}{\sigma^2(1-F(u_{1t}))^2} + \frac{u_{1t} f(u_{1t})[(u_{1t}/\sigma)^2 - 1]}{\sigma^2(1-F(u_{1t}))} \right\}$$

$$- n^{-1} \sum_{S_2} \left\{ \frac{u_{2t}^2 f^2(u_{2t})}{\sigma^2 F^2(u_{2t})} - \frac{u_{2t} f(u_{2t})[(u_{2t}/\sigma)^2 - 1]}{\sigma^2 F(u_{2t})} \right\}$$

$$- n^{-1} \sum_{S_3} \left\{ \{\sigma^{-2} - 3\sigma^{-4}(y_t - g(\underline{x}_t; \underline{\theta}))^2\}. \right\}$$

Assuming that the exogenous variables are nonrandom, the randomness in the summations of the first two terms of 5.5 is due to the random sets S_1 and S_2. As before, the *summation over* S_1 may be replaced by a summation over t from 1 to n, if the indicator function $I(z_t < L_{1t})$ is written as a factor of each of the terms of the summation. We have that

$$E(I(z_t < L_{1t})) = 1-F(u_{1t})$$

$$E(I(z_t > L_{2t})) = F(u_{2t})$$

and the expectations of the first two terms of 5.4, 5.5, and 5.6 can be computed. To invoke theorems 1.1, 1.2, and 1.3, each of these expectations must converge as $n \to \infty$, uniformly for $\underline{\theta},\sigma^2$ in a compact

set. To compute the expectations of the third terms of 5.4, 5.5, and 5.6, we note that

$$E(I(L_{1t} \leq z_t \leq L_{2t})) = F(u_{1t}) - F(u_{2t})$$

$$E((z_t - g(\underline{x}_t; \underline{\theta}))I(L_{1t} \leq z_t \leq L_{2t})) = \int_{u_{2t}}^{u_{1t}} \varepsilon f(\varepsilon) d\varepsilon .$$

It must be assumed that the expectations of the third terms converge in the sense specified previously. In order to show that the variances of 5.4, 5.5, and 5.6 converge to zero as required, it is necessary to assume that g and g_i', for $i=1,\ldots,K$, are bounded in absolute value uniformly for all $\underline{x}_t$, that g_{ij}'', for $i,j=1,\ldots,K$, is bounded uniformly in absolute value for all $\underline{x}_t$ and $\underline{\theta}$ in any compact set, and the sequences $\{L_{1t}\}$ and $\{L_{2t}\}$ are bounded uniformly in t.

5.4 A Threshold Regression Model

The threshold regression model of Dagenais [1975] specifies that the value of the dependent variable y_t remains fixed at a value L until the action of either of two groups of exogenous variables forces it across either a lower or an upper threshold value. Unlike the two-limit probit model, the dependent variable of the threshold model has neither upper nor lower bounds. The behavior of y_t is described

below. Let

$$y_t = z_{1t} \quad \text{if} \quad z_{1t} < L_{1t}$$

$$y_t = L \quad \text{if} \quad z_{1t} \geq L_{1t},\ z_{2t} \leq L_{2t}$$

$$y_t = z_{2t} \quad \text{if} \quad z_{2t} > L_{2t}$$

where

$$z_{1t} = \beta_1 + \underline{\theta}^T \underline{x}_t - u_t$$

$$z_{2t} = \beta_2 + \underline{\theta}^T \underline{x}_t - u_t \qquad \beta_1 \geq \beta_2$$

$$L_{1t} = L - \underline{\xi}^T \underline{R}_t + \varepsilon_{1t} \leq L$$

$$L_{2t} = L + \underline{\zeta}^T \underline{Q}_t + \varepsilon_{2t} \geq L.$$

The variables $\underline{x}_t$, $\underline{R}_t$, and $\underline{Q}_t$ are exogenous. The variables $\{\underline{x}_t,\ t=1,\ldots\}$ determine the regression hyperplane. The variables $\{\underline{R}_t,\ t=1,\ldots\}$ determine the lower threshold. The variables $\{\underline{Q}_t,\ t=1,\ldots\}$ determine the upper threshold. The variables u_t, ε_{1t}, and ε_{2t} are error terms. The constant L is assumed known. The unknown parameters are β_1, β_2, $\underline{\theta}$, $\underline{\xi}$, $\underline{\zeta}$. and the parameters of the distribution of the error terms.

For the sake of convenience, we will assume the error variables u_t, ε_{1t}, and ε_{2t} are distributed independently of each other, and are serially independent over time. Let F_i, f_i be the c.d.f. and density, respectively, of a normal variate with expectation zero and

variance σ_i^2, for $i=0, 1,$ and 2. Then the joint density of u_t, ε_{1t}, and ε_{2t} is

$$\begin{array}{lll}
f_0(u_t)f_1(\varepsilon_{1t})f_2(\varepsilon_{2t}) & \text{if} & \varepsilon_{1t} < \underline{\xi}^T\underline{R}_t,\ \varepsilon_{2t} > -\underline{\zeta}^T\underline{Q}_t \\
f_0(u_t)(1-F(\underline{\xi}^T\underline{R}_t))f_2(\varepsilon_{2t}) & \text{if} & \varepsilon_{1t} = \underline{\xi}^T\underline{R}_t,\ \varepsilon_{2t} > -\underline{\zeta}^T\underline{Q}_t \\
f_0(u_t)f_1(\varepsilon_{1t})F_2(-\underline{\zeta}^T\underline{Q}_t) & \text{if} & \varepsilon_{1t} < \underline{\xi}^T\underline{R}_t,\ \varepsilon_{2t} = -\underline{\zeta}^T\underline{Q}_t \\
f_0(u_t)(1-F_1(\underline{\xi}^T\underline{R}_t))F_2(-\underline{\zeta}^T\underline{Q}_t) & \text{if} & \varepsilon_{1t} = \underline{\xi}^T\underline{R}_t,\ \varepsilon_{2t} = -\underline{\zeta}^T\underline{Q}_t.
\end{array}$$

Let $S_1 = \{t:\ z_{1t} < L_{1t}\}$, $S_2 = \{t:\ z_{1t} \geq L_{1t},\ z_{2t} \leq L_{2t}\}$, $S_3 = \{t:\ z_{2t} \geq L_{2t}\}$. The logarithm of the joint density of $y_1, \ldots, y_n$ is

$$\log L_n(\beta_1,\ \beta_2,\ \underline{\theta},\ \underline{\xi},\ \underline{\zeta},\ \sigma_0^2,\ \sigma_1^2,\ \sigma_2^2) =$$

$$\sum_{S_1} \log[f_0(\beta_1 + \underline{\theta}^T\underline{x}_t - y_t)(1-F_1(y_t - L + \underline{\xi}^T\underline{R}_t))]$$

$$+ \sum_{S_2} \log[\int_{-\infty}^{G_2} f_0(u)(1-F_2(\beta_2 + \underline{\theta}^T\underline{x}_t - L - u - \underline{\zeta}^T\underline{Q}_t))du$$

$$+ F_0(\beta_1 + \underline{\theta}^T\underline{x}_t - L) - F_0(\beta_2 + \underline{\theta}^T\underline{x}_t - L)$$

$$+ \int_{G_1}^{\infty} f_0(u)F_1(\beta_1 + \underline{\theta}^T\underline{x}_t - L - u + \underline{\xi}^T\underline{R}_t)du]$$

$$+ \sum_{S_3} \log[f_0(\beta_2 + \underline{\theta}^T\underline{x}_t - y_t)F_2(y_t - L - \underline{\zeta}^T\underline{Q}_t)]$$

where $G_2 = \beta_2 + \underline{\theta}^T\underline{x}_t - L$ and $G_1 = \beta_1 + \underline{\theta}^T\underline{x}_t - L$. Let $\underline{B}^T =$

$(\beta_1, \beta_2, \underline{\theta}^T, \underline{\xi}^T, \underline{\zeta}^T)$ and let $\underline{V}_t^T$ be the corresponding vector of variables $(1, 1, \underline{x}_t^T, \underline{R}_t^T, \underline{Q}_t^T)$. The second partial derivatives of the log likelihood function may be expressed as

$$- n^{-1} \frac{\partial^2 \log L_n(B, \sigma_0^2, \sigma_1^2, \sigma_2^2)}{\partial B_i \partial B_j} = - n^{-1} \sum_{t=1}^{n} h_{ij}(t)(\underline{V}_t)_i(\underline{V}_t)_j$$

where $h_{ij}(t)$ is a continuous function of B, σ_0^2, σ_1^2, σ_2^2, and $\underline{V}_t$. We must assume that the above term converges as $n \to \infty$, uniformly when the parameters are in any compact set.

We can express the second partials with respect to the covariance terms as

5.7 $$- n^{-1} \frac{\partial^2 \log L_n(B, \sigma_0^2, \sigma_1^2, \sigma_2^2)}{\partial^2 \sigma_0} = - n^{-1} \sum_{S_1} [\sigma_0^{-2} - 3(\beta_1 + \underline{\theta}^T \underline{x}_t - y_t)^2 \sigma_0^{-4}]$$

$$- n^{-1} \sum_{S_2} [\text{nonrandom function of } t]$$

$$- n^{-1} \sum_{S_3} [\sigma_0^{-2} - 3(\beta_2 - \underline{\theta}^T \underline{x}_t - y_t)^2 \sigma_0^{-4}].$$

5.8 $$- n^{-1} \frac{\partial^2 \log L_n(B, \sigma_0^2, \sigma_1^2, \sigma_2^2)}{\partial^2 \sigma_1} =$$

$$- n^{-1} \sum_{S_1} \frac{(y_t - L + \underline{\xi}^T \underline{R}_t)^2 f_1^2(y_t - L + \underline{\xi}^T \underline{R}_t)}{\sigma_1^2 (1 - F_1(y_t - L + \underline{\xi}^T \underline{R}_t))^2}$$

$$- n^{-1} \sum_{S_1} \frac{(y_t - L + \underline{\xi}^T \underline{R}_t) f_1(y_t - L + \underline{\xi}^T \underline{R}_t)}{\sigma_1^2 (1 - F_1(y_t - L + \underline{\xi}^T \underline{R}_t))} \left\{ \frac{(y_t - L + \underline{\xi}^T \underline{R}_t)^2}{\sigma_1^2} - 1 \right\}$$

$$- n^{-1} \sum_{S_2} [\text{nonrandom function of } t].$$

5.9 $$- n^{-1} \frac{\partial^2 \log L_n(\underline{B}, \sigma_0^2, \sigma_1^2, \sigma_2^2)}{\partial^2 \sigma_2} =$$

$$- n^{-1} \sum_{S_2} [\text{nonrandom function of } t]$$

$$- n^{-1} \sum_{S_3} \frac{(y_t - L - \underline{\xi}^T \underline{Q}_t)^2 f_2^2 (y_t - L - \underline{\xi}^T \underline{Q}_t)}{\sigma_2^2 F_2^2 (y_t - L - \underline{\xi}^T \underline{Q}_t)}$$

$$+ n^{-1} \sum_{S_3} \frac{(y_t - L - \underline{\xi}^T \underline{Q}_t) f_2(y_t - L - \underline{\xi}^T \underline{Q}_t)}{\sigma_2^2 F_2(y_t - L - \underline{\xi}^T \underline{Q}_t)} \left\{ \frac{(y_t - L - \underline{\xi}^T \underline{Q}_t)^2}{\sigma_2^2} - 1 \right\}$$

The cross derivatives with respect to the variance terms are either nonrandom or zero. The expectations of 5.7, 5.8, and 5.9 must be assumed to converge as $n \to \infty$ uniformly for the parameters in any compact set. Since the error terms were assumed to be serially independent, the variances of the second partials converge as required when the exogenous variables are contained in a compact set.

5.5 A Simultaneous Logit Model

Schmidt and Strauss [1975] determine the MLE of the parameters of a simultaneous version of a multivariate logit model. The variable

x_t is assumed to have I possible values: $1,\ldots,I$. The variable y_t is assumed to have J possible values: $1,\ldots,J$. Let

$$x_{kt}^* = 1 \quad \text{if} \quad x_t = k, \quad \text{for} \quad k=2,\ldots,I$$

$$x_{kt}^* = 0 \quad \text{otherwise.}$$

$$y_{kt}^* = 1 \quad \text{if} \quad y_t = k, \quad \text{for} \quad k=2,\ldots,J$$

$$y_{kt}^* = 0 \quad \text{otherwise.}$$

The joint distribution of $\{x_t, t=1,\ldots\}$ and $\{y_t, t=1,\ldots\}$ is given by the model

$$\log\left\{\frac{P(x_t = i/y_t)}{P(x_t = 1/y_t)}\right\} = \underline{\beta}_i^T\underline{R}_t + \sum_{k=2}^{J} \alpha_{ik} y_{kt}^* \qquad i=2,\ldots,I$$

$$\log\left\{\frac{P(y_t = j/x_t)}{P(y_t = 1/x_t)}\right\} = \underline{\gamma}_j^T\underline{Q}_t + \sum_{k=2}^{I} \alpha_{kj} x_{kt}^* \qquad j=2,\ldots,J.$$

$\underline{R}_t$ is a Kx1 vector; $\underline{Q}_t$ is a Lx1 vector. Again, let $I(E)$ represent the indicator function of the event E. Let

$$e_t = 1 + \sum_{i=2}^{I} \exp(\underline{\beta}_i^T\underline{R}_t) + \sum_{j=2}^{J} \exp(\underline{\gamma}_j^T\underline{Q}_t)$$

$$+ \sum_{i=2}^{I}\sum_{j=2}^{J} \exp(\underline{\beta}_i^T\underline{R}_t + \underline{\gamma}_j^T\underline{Q}_t + \alpha_{ij})$$

Then the likelihood function of the observations is

$$\sum_{t=1}^{n} - \log(e_t)$$

$$+ \sum_{t=1}^{n} \sum_{j=2}^{J} I(x_t = 1,\ y_t = j)(\underline{\gamma}_j^T \underline{Q}_t)$$

$$+ \sum_{t=1}^{n} \sum_{i=2}^{I} I(x_t = i,\ y_t = 1)(\underline{\beta}_i^T \underline{R}_t)$$

$$+ \sum_{t=1}^{n} \sum_{i=2}^{I} \sum_{j=2}^{J} I(x_t = i,\ y_t = j)(\underline{\beta}_i^T \underline{R}_t + \underline{\gamma}_j^T \underline{Q}_t + \alpha_{ij})$$

In order to compactly express the second partials, the following notation will be convenient.

$$z_{tab} = e_t^{-1} \exp(\underline{\beta}_a^T \underline{R}_t) R_{tb} [1 + \sum_{j=2}^{J} \exp(\underline{\gamma}_j^T \underline{Q}_t + \alpha_{aj})] \quad \text{for} \quad a=2,\ldots,I,\ b=1,\ldots,K$$

$$w_{tcd} = e_t^{-1} \exp(\underline{\gamma}_c^T \underline{Q}_t) Q_{td} [1 + \sum_{i=2}^{I} \exp(\underline{\beta}_i^T \underline{R}_t + \alpha_{ic})] \quad \text{for} \quad c=2,\ldots,J,\ d=1,\ldots,L.$$

$$s_{tij} = e_t^{-1} \exp(\underline{\beta}_i^T \underline{R}_t + \underline{\gamma}_j^T \underline{Q}_t + \alpha_{ij}) \quad \text{for} \quad i=2,\ldots,I,\ j=2,\ldots,J.$$

Let β_{ab}, γ_{ab} be the b'th elements of $\underline{\beta}_a$, $\underline{\gamma}_a$ respectively. Let $\delta_{hk} = 1$ if $h=k$, and 0 otherwise. Then the second partial derivatives can be expressed as follows.

5.10 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\beta_{ab}\partial\beta_{cd}} = - n^{-1} \sum_{t=1}^{n} z_{tab}(z_{tcd} - \delta_{ac}R_{td})$$

for $a,c=2,\ldots,I;\ b,d=1,\ldots,K.$

5.11 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\gamma_{ab}\partial\gamma_{cd}} = - n^{-1} \sum_{t=1}^{n} w_{tab}(w_{tcd} - \delta_{ac}Q_{td})$$

for $a,c=2,\ldots,J;\ b,d=1,\ldots,L.$

5.12 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\beta_{ab}\partial\gamma_{cd}} = - n^{-1} \sum_{t=1}^{n} z_{tab}w_{tcd} + n^{-1} \sum_{t=1}^{n} R_{tb}Q_{td}s_{tac}$$

for $a=2,\ldots,I;\ b=1,\ldots,K;$

$c=2,\ldots,J;\ d=1,\ldots,L.$

5.13 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\alpha_{ab}\partial\alpha_{cd}} = - n^{-1} \sum_{t=1}^{n} s_{tab}(s_{tcd} - \delta_{ac}\delta_{bd})$$

for $a,c=2,\ldots,I;\ b,d=2,\ldots,J.$

5.14 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\beta_{ab}\partial\alpha_{cd}} = - n^{-1} \sum_{t=1}^{n} s_{tcd}(z_{tab} - \delta_{ac}R_{tb})$$

for $a,c=2,\ldots,I;\ b=1,\ldots,K;\ d=2,\ldots,J.$

5.15 $$- n^{-1} \frac{\partial^2 \log L_n}{\partial\gamma_{ab}\partial\alpha_{cd}} = - n^{-1} \sum_{t=1}^{n} s_{tcd}(w_{tab} - \delta_{ad}Q_{tb})$$

for $a,d=2,\ldots,J;\ b=1,\ldots,L;\ c=2,\ldots,I.$

If the exogenous variables $\underline{R}_t$ and $\underline{Q}_t$ are nonrandom, the convergence of terms 5.10 to 5.15 uniformly in compact sets of the unknown parameters, is sufficient to assure that the MLE are consistent, asymptotically normally distributed, and efficient in the MP sense.

BIBLIOGRAPHY

Amemiya, Takeshi, and Boskin, Michael. [1974]. "Regression Analysis when the Dependent Variable is Truncated Lognormal, with an Application to the Determinants of the Duration of Welfare Dependency." Inter. Econ. Review, vol. 15, no. 2, pp. 485-496.

Anderson, T.W., and Rubin, H. [1949]. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." Ann. Math. Stat., vol. 20, no. 1, pp. 46-63.

Anderson, T.W., and Rubin, H. [1950]. "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations." Ann. Math. Stat., vol. 21, no. 4, pp. 570-582.

Barnett, William A. [1974]. "Maximum Likelihood Estimation of Nonlinear Systems of Equations." Division of Research and Statistics, Federal Reserve Board, Special Studies Paper, Washington, D.C.

Bar-Shalom, Yaakov. [1971]. "On the Optimal Properties of the Maximum Likelihood Estimate Obtained from Dependent Observations." Jour. Royal Stat. Soc., Ser. B, vol. 33, no. 1, pp. 72-77.

Barten, A.P. [1969]. "Maximum Likelihood Estimation of a Complete System of Demand Equations." Eur. Econ. Review, vol. 1, no. 1, pp. 7-73.

Berndt, E.R., and Savin, N.E. [1975]. "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances." Econometrica, vol. 43, no. 5, pp. 937-957.

Bhat, B.R. [1974]. "On the Method of Maximum Likelihood for Dependent Observations." Jour. Royal Stat. Soc., Ser. B, vol. 36, no. 1, pp. 48-53.

Cramér, Harald. [1946]. "Mathematical Methods of Statistics". Princeton University Press, Princeton.

Crowder, Martin J. [1976]. "Maximum Likelihood Estimation for Dependent Observations." Jour. Royal Stat. Soc., ser. B, vol. 38, no. 1, pp. 45-53.

Dagenais, M.G. [1975]. "Application of a Threshold Regression Model to Household Purchases of Automobiles." Review of Econ. and Stat., vol. 57, no. 3, pp. 275-285.

Durbin, J. [1961]. "Estimation of Parameters in Time Series Regression Models." Jour. Royal Stat. Soc., ser. B, vol. 22, no. 1, pp. 139-153.

Goldberger, A.S. [1964]. "Econometric Theory." John Wiley and Sons, New York.

Grenander, Ulf, and Rosenblatt, Murray. [1957]. "Statistical Analysis of Stationary Time Series." John Wiley and Sons, New York.

Guilkey, D. and Schmidt, P. [1973]. "Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Disturbances." Jour. Amer. Stat. Assoc., vol. 68, no. 343, pp. 642-647.

Hannan, E.J. [1971]. "Nonlinear Time Series Regression." Jour. Appl. Prob., vol. 8, no. 4, pp. 767-780.

Hendry, D.F. [1971]. "Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process." Inter. Econ. Review, vol. 12, no. 2, pp. 257-272.

Jennrich, R.I. [1969]. "Asymptotic Properties of Nonlinear Least Squares Estimators." Ann. Math. Stat., vol. 40, no. 2, pp. 633-643.

Jöreskog, K.G., and Goldberger, A.S. [1975]. "Estimators of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable." Jour. Amer. Stat. Assoc., vol. 70, no. 351, pp. 631-639.

Kmenta, Jan, and Gilbert, R.F. [1970]. "Estimation of Seemingly Unrelated Regressions with Autoregressive Disturbances." Jour. Amer. Stat. Assoc., vol. 65, no. 329, pp. 186-197.

Koopmans, T.C., and Hood, W.C. [1953]. "The Estimation of Simultaneous Linear Economic Relationships." Cowles Commission for Research in Economics, Monograph 14, Chap. VI, John Wiley and Sons, New York.

Koopmans, T.C., Rubin, H. and Leipnik, R.B. [1950]. "Measuring the Equation Systems of Dynamic Economics." Cowles Commission for Research in Economics, Monograph 10, Chap. II, John Wiley and Sons, New York.

Malinvaud, Edmond. [1966]. "Statistical Methods of Econometrics." North Holland Publishing Company, Amsterdam.

Malinvaud, Edmond. [1970]. "The Consistency of Nonlinear Regressions." Ann. Math. Stat., vol. 41, no. 3, pp. 956-969.

Mann, H.B., and Wald, A. [1943]. "On the Statistical Treatment of Linear Stochastic Difference Equations.", Econometrica, vol. 11, no. 3, pp. 173-220.

Nocturne, D.J. [1970]. "Asymptotic Efficiency of the Maximum Likelihood Estimators for the Parameters of Certain Stochastic Processes." Unpublished thesis, Cornell University.

Parks, R.W. [1967]. "Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated." Jour. Amer. Stat. Assoc., vol. 62, no. 318, pp. 500-509.

Robinson, P.M. [1972]. "Nonlinear Regression for Multiple Time Series." Jour. Appl. Prob., vol. 9, no. 4, pp. 758-768.

Rosett, R.N., and Nelson, F.D. [1975]. "Estimation of the Two-Limit Probit Regression Model." Econometrica, vol. 43, no. 1, pp. 141-146.

Rubin, H. and Chernoff, H. [1953]. "Asymptotic Properties of Limited Information Estimates Under Generalized Conditions." Cowles Commission for Research in Economics, monograph 14, chap. VII, John Wiley and Sons, New York.

Schmidt, Peter, and Strauss, R.P. [1975]. "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach." Econometrica, vol. 43, no. 4, pp. 745-755.

Silvey, S.D. [1961]. "A Note on Maximum Likelihood in the Case of Dependent Random Variables." Jour. Royal Stat. Soc., ser. B, vol. 23, no. 2, pp. 444-452.

Theil, Henri. [1971]. "Principles of Econometrics." John Wiley and Sons, New York.

Tobin, J. [1958]. "Estimation of Relationships for Limited Dependent Variables." Econometrica, vol. 26, no. 1, pp. 24-36.

Wald, A. [1948]. "Asymptotic Properties of the Maximum Likelihood Estimate of an Unknown Parameter of a Discrete Stochastic Process." Ann. Math. Stat., vol. 19, no. 1, pp. 40-46.

Wald, A. [1949]. "Note on the Consistency of the Maximum Likelihood Estimate." Ann. Math. Stat., vol. 20, no. 4, pp. 595-601.

Weiss, L.I. [1971]. "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases." Jour. Amer. Stat. Assoc., vol. 66, no. 334, pp. 345-350.

Weiss, L.I. [1973]. "Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases, II." Jour. Amer. Stat. Assoc., vol. 68, no. 342, pp. 428-430.

Weiss, L.I. and Wolfowitz, J. [1967]. "Maximum Probability Estimators." Ann. Inst. Stat. Math., vol. 19, no. 2, pp. 193-206.

1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

OPTIMAL ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS OF PARAMETERS OF SOME ECONOMETRIC MODELS

5. TYPE OF REPORT & PERIOD COVERED

Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Mary Kathleen Vickers

8. CONTRACT OR GRANT NUMBER(s)

DAA G29-75-C-0008

N00014-75-C-0586

9. PERFORMING ORGANIZATION NAME AND ADDRESS

School of Operations Research & Industrial Engineering, College of Engineering
Cornell University, Ithaca, NY 14853

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Sponsoring Military Activity
U.S. Army Research Office, Post Office Box 12211
Research Tringle Park, NC 27709

12. REPORT DATE

May 1977

13. NUMBER OF PAGES

140

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

Sponsoring Military Activity
Statistics and Probability Program
Office of Naval Research
Arlington, VA 22217

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

NA

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an offical Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonlinear Models
Efficient Estimation
Econometric models
Maximum Likelihood Estimation
Autoregressive Models
Probit and Logit Models
Systems of Equations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Four theorems are proven, which simplify the application to econometric models of Weiss's theorem ~~[1973]~~ on asymptotic properties of maximum likelihood estimators in nonstandard cases. The theorems require, roughly: the uniform convergence in any compact sets of the unknown parameters of the expection of the Hessian matrix of the log likelihood function; and the uniform convergence to 0 in the same sense of the variance of the same quantities. The fourth theorem allows one to conclude that the optimal properties hold on an image → next page

(cont'd on back)

set of the parameters when the map satisfies certain smoothness conditions, and the first three theorems are satisfied for the original parameter set. These four theorems are applied to autoregressive models, nonlinear models, systems of equations, and probit and logit models to infer optimal asymptotic properties.